

th J, Miller RH, Purcell RH;
1; 96-139709/14.
2:DB; T16650.
3: and amino acid sequence of HCV envelope 1 and core proteins -
4: to determine HCV genotype and as vaccines against HCV infection
5: m 4; Page 214; 340pp; English.
6: 936-R92987 are HCV core proteins derived from 52 different HCV
7: plates. Isolated cDNA sequences are used for the prodn. of primers
8: for detecting the presence of HCV in a sample, the primers
9: are also useful for HCV genotyping. Proteins encoded by the cDNAs
10: are used in vaccines for immunising against HCV infection. The
11: proteins may also be used to detect antibodies against HCV in serum,
12: saliva, lymphocytes or other mononuclear cells. The antibodies may
13: be used in the prevention of HCV infection.
14: pence 191 AA;

Match 99.5%; Score 777; DB 17; Length 191;
Local Similarity 90.8%; Pred. No. 2.12e-61;
9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
1 matnpgqtktrntnrrpdxkfgggqivggvllprgrprgvrtrtktsersqrg 60
1 MSTNPKQKTKNTNRRPXXXXPFGGQIVGGVLLPRGRPRGXVTRTKTSERSQPRG 60
1 rrpdpkarrsegrswagqypwplynecgagwllprgrprswq 109
1 RRPDPKARRSEGRSWAGQYPPWPLYNECGAGWLLSPRGRPNWGP 109

4
373 standard; Protein; 191 AA.
373;
ACT-1996 (first entry)
itis C virus isolate 28 core protein.
E; envelope 1; core protein; HCV genotyping; antibody; vaccine;
itis C virus.
-05315-AZ.
EB-1996.
UG-1995; U10398.
UG-1994; US-290665.
H) US DEPT HEALTH & HUMAN SERVICES.
H) US SEC DEPT HEALTH.
J, Miller RH, Purcell RH;
96-139709/14.
DB; T16647.
and amino acid sequence of HCV envelope 1 and core proteins -
to determine HCV genotype and as vaccines against HCV infection
m 4; Page 211-212; 340pp; English.
35-R92987 are HCV core proteins derived from 52 different HCV
plates. Isolated cDNA sequences are used for the prodn. of primers
for detecting the presence of HCV in a sample, the primers
are also useful for HCV genotyping. Proteins encoded by the cDNAs
are used in vaccines for immunising against HCV infection. The
proteins may also be used to detect antibodies against HCV in serum,
saliva, lymphocytes or other mononuclear cells. The antibodies may
be used in the prevention of HCV infection.
ence 191 AA;

Match 99.5%; Score 777; DB 17; Length 191;
Local Similarity 90.8%; Pred. No. 2.12e-61;
9; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
matnpgqtktrntnrrpdxkfgggqivggvllprgrprgvrtrtktsersqrg 60

Oy 1 MSTNPKQKTKNTNRRPXXXXPFGGQIVGGVLLPRGRPRGXVTRTKTSERSQPRG 60
Db 61 rrpdpkarrsegrswagqypwplynecgagwllprgrprswq 109
Oy 61 RRPDPKARRSEGRSWAGQYPPWPLYNECGAGWLLSPRGRPNWGP 109

RESULT 5
ID R67591 standard; Protein; 502 AA.
AC R67591;
DE 07-SEP-1995 (first entry)
DE Hepatitis C virus SR037-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; SR037-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI; 95-040318/06.
DR N-PSDB; 079143.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
treatment of hepatitis C
PS Claim 21; Page 36; 41pp; Japanese.
CC 079143 is the hepatitis C virus (HCV) gene SR037-5' cDNA, it encodes
the protein described in R67591. Both the cDNA and protein can be
used in the treatment of HCV infection
SQ Sequence 502 AA;

Query Match 99.5%; Score 777; DB 12; Length 502;
Best Local Similarity 90.8%; Pred. No. 2.12e-61;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 matnpgqtktrntnrrpdxkfgggqivggvllprgrprgvrtrtktsersqrg 60
Oy 1 MSTNPKQKTKNTNRRPXXXXPFGGQIVGGVLLPRGRPRGXVTRTKTSERSQPRG 60
Db 61 rrpdpkarrsegrswagqypwplynecgagwllprgrprswq 109
Oy 61 RRPDPKARRSEGRSWAGQYPPWPLYNECGAGWLLSPRGRPNWGP 109

RESULT 6
ID R67589 standard; Protein; 502 AA.
AC R67589;
DE 07-SEP-1995 (first entry)
DE Hepatitis C virus YS117-5' gene product.
KW Hepatitis C virus; HCV; non-A non-B; YS117-5'; treatment.
OS Hepatitis C virus.
PN J06319563-A.
PD 22-NOV-1994.
PF 13-MAY-1993; 147133.
PR 13-MAY-1993; JP-147133.
PA (IMMO) IMMUNO JAPAN KK.
DR WPI; 95-040318/06.
DR N-PSDB; 079141.
PT A hepatitis C virus gene and oligo-nucleotide(s) - used for the
treatment of hepatitis C
PS Claim 19; Page 34; 41pp; Japanese.
CC 079141 is the hepatitis C virus (HCV) gene YS117-3' cDNA, it encodes
the protein described in R67588. Both the cDNA and protein can be
used in the treatment of HCV infection
SQ Sequence 502 AA;


```

RESEQUENCE ID A50346 standard; DNA; UNC; 327 BP.
AC A50346;
WI e1039185
OT 07-MAR-1997 (Rel. 51, Created)
OT 07-MAR-1997 (Rel. 51, Last updated, Version 1)
DE Sequence 1 from Patent W09613590.
KW .
UN unidentified
UN unclassified.
WI [1]
WI 1-327
MA Maertens G., Stuyver L.;
AB "NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE
AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS";
PI Patent number W09613590-A/1, 09-MAY-1996.
TI INNOGENETICS NV (BE).
AU Other publication AU 3844095 960523
CH Key Location/Qualifiers
H source 1..327
T /organism="unidentified"
Q Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;

Query Match 100.0%; Score 317; DB 143; Length 327;
Best Local Similarity 100.0%; Pred. No. 7.68e-244;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps

bb 1 atgacgacgaatcctaactcaagaagaaccaaactaacacacgcgcgcctcaak 60
yy 1 ATGACGACGAATCCTAAACCTCAAGAAACAAACGTAACACCAACCGCGCCCTCAK 60
bb 61 gsgtnnnnnccgggtgcgcgcgactaggaagactccgagcggtcacacactcgtggc 120
yy 61 GSGTNNNNNCCGGGTGCGCGGCTCAGATCCTTGGTGAGCTTTACTCTGTCCGCGCAGG 120
b 121 ggcacagcagcagcgtgtgcgcgcgactaggaagactccgagcggtcacacactcgtggc 180
yy 121 GGCACAGCAGCGGTGTGCGCGGCGACTAGGAAGACTTCCGAGCGGTCACAACTGTGTGGC 180
b 181 aggcacagcagcctatccccaaagcctgcgcgcgcagggcaggtcctggcctcagcccg 240
yy 181 AGGCACAGCCTATCCCCAAGGCTGCGCGGTCCGAGGGCGAGGTCCTGGGCTCAGCCGGG 240
b 241 tatcctggccctctatggcaatgaggcctgcgggtggcgagggtggcctcgtgcctccccc 300
yy 241 TATCCTGGCCCTCTATGGCAATGAGGCGTGGCGGTGGCGGCTGGGCTCAGCCGCC 300
b 301 cgcgcctctcgcgcacaaattggggcccc 327
yy 301 CGCGGCTCTCGGCCCAATTGGGCCCC 327

```

RESULT	2	HCVN3	506 bp	DNA	VRL
OCUS					
DEFINITION		Hepatitis C virus core gene, HN3 isolate.			
ACCESSION		X76410			
LOCUS		943497			
KEYWORDS		core; core protein.			
SOURCE		Hepatitis C virus.			
ORGANISM		Hepatitis C virus			
		Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.			

REFERENCE 1 (bases 1 to 506)
AUTHORS Qu, D.
TITLE Direct Submission
JOURNAL Submitted (22-NOV-1993) to the EMBL/GenBank/DBJ databases. D. Qu,
INSERM Unite 271, U. de Recherche sur les Hepatites, le Sida et
les Retrovirus Humains, 151 cours Albert Thomas, F 69424 Lyon Cedex
03, FRANCE

REFERENCE 2 (bases 1 to 506)
AUTHORS Qu, D., Hantz, O., Gouy, M., Vitvitski, L., Li, J.S., Berby, F.,
Tong, S.P. and Trepo, C.
TITLE Heterogeneity of hepatitis C virus genotypes in France
J. Gen. Virol. 75 (Pt 5), 1063-1070 (1994)
MEDLINE 94231157
FEATURES Location/Qualifiers
source 1..506
/organism="Hepatitis C virus"
/isolate="HN3"
misc_feature 29..506
/notes="region corresponding to the core protein of the
polyprotein of HCV"

BASE COUNT 86 a 166 c 166 g 88 t
ORIGIN

Query Match 91.8%; Score 291; DB 117; Length 506;
Best Local Similarity 91.7%; Pred. No. 4.45e-221;
Matches 300; Conservative 4; Mismatches 23; Indels 0; Gaps 0;

Ddb 29 atgagcagaatcctaaacctcaaggaaacccaaacgttaacatcaacgcgcgcacag 88
Qy 1 ATGACAGCATCTTAACCTCAAGAAACAAACGCTACACACACCGCGCCCTCAK 60
Ddb 89 gacgtcaagttcccgggcggtggtcagatcgttggtggagttacctgttgcgcgcag 148
Qy 61 GCGGTNNNNNNCCGGTGGCGGCTGAGATCGTGTGTGGAGATTACCTGTTCCGCGCAGG 120
Ddb 149 gcccacagttgggtgtgcgcgcactaggaagacttcgcagagctgcgaacctcgtgac 208
Qy 121 GCGCCCGCAGGNGGGTGTGGCGCGCATAGCAAGACTTCCGACGCGCTCACACCTCGTGC 180
Ddb 209 aggcgcagcgtatccccaaaggtcgtgcgcgcagggcgaggtcctggctcagccggg 268
Qy 181 AGGCGACAGCCTATCCCAAGGCTGCGGCGCAGGCAAGTCTCGGCTCAGCCCGGG 240
Ddb 269 tacccttgccctctatggcaacagggtcgcggtggcagaggtggctcgtgcctc 328
Qy 241 TATCCTTGCCCTCTATGGCAGTAGGCGTGGCGGTGGCGGGGWTGGCTGCTGTCCCCC 300
Ddb 329 cgtggtctcgccctagttggggcccc 355
Qy 301 CGCGGCTCTCGGCCCAATTGGGGGCCCC 327

RESULT 3
LOCUS HCUI0204 573 bp RNA VRL 26-AUG-1994
DEFINITION Hepatitis C virus isolate P10 core protein gene
ACCESSION U10204
KEYWORDS g532392
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (bases 1 to 506)
AUTHORS Buhk, J., Purcell, R.H. and Miller, R.H.

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	777	99.5	191	13	PCT-US95-1	Sequence 195, Applicant	2.00e-59
2	777	99.5	191	13	PCT-US95-1	Sequence 192, Applicant	2.00e-59
3	775	99.2	191	13	PCT-US95-1	Sequence 158, Applicant	3.03e-59
4	775	99.2	191	13	PCT-US95-1	Sequence 156, Applicant	3.03e-59
5	775	99.2	191	13	PCT-US95-1	Sequence 157, Applicant	3.03e-59
6	775	99.2	191	13	PCT-US95-1	Sequence 196, Applicant	3.03e-59
7	775	99.2	191	13	PCT-US95-1	Sequence 159, Applicant	3.03e-59
8	775	99.2	191	13	PCT-US95-1	Sequence 160, Applicant	3.03e-59
9	775	99.2	967	12	PCT-US94-0	Sequence 13, Applicant	3.03e-59
10	775	99.2	967	13	PCT-US95-0	Sequence 13, Applicant	3.03e-59
11	775	99.2	967	7	US-08-188-	Sequence 13, Applicant	3.03e-59
12	775	99.2	1648	13	PCT-US95-0	Sequence 12, Applicant	3.03e-59
13	775	99.2	1648	12	PCT-US94-0	Sequence 12, Applicant	3.03e-59

GENERAL INFORMATION. CC
CC APPLICANT: RUKH, J., MILLER, R.H. AND CC
CC APPLICANT: PORCELLI, R.H. CC
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED CC
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND CC
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS CC
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE CC
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES CC
CC NUMBER OF SEQUENCES: 263 CC
CC CORRESPONDENCE ADDRESS: CC
CC ADDRESSEE: MORGAN & FINNEGAN CC
CC STREET: 345 PARK AVENUE CC
CC CITY: NEW YORK CC
CC STATE: NEW YORK CC
CC COUNTRY: USA CC
CC ZIP: 10154 CC
CC COMPUTER READABLE FORM: CC
CC MEDIUM TYPE: FLOPPY DISK CC

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

3

CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 195:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: 26
SQ SEQUENCE 191 AA; 20774 MW; 184617 CN;

Query Match . . . 99.5%; Score 777; DB i3; Length 191;
Best Local Similarity 90.8%; Pred. No. 2.00e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKNTNRRPMDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKNTNRRPMDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
|||||

Db 61 RRQIPKARSEGRSWAQPYPMPLYGNECGWAGWLLSPRGRSPWGP 109
|||||
Qy 61 RRQIPKARSEGRSWAQPYPMPLYGNECGWAGWLLSPRGRSPWGP 109
|||||

RESULT 2
ID PCT-US95-10398-192 STANDARD; PRT; 191 AA.
XX AC xxxxxx
XX 01-JAN-1900
XX
DE Sequence 192, Application PC/TUS9510398.
XX
CC Sequence 192, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

4

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 192:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: 28
SQ SEQUENCE 191 AA; 20744 MW; 181593 CN;

Query Match . . . 99.5%; Score 777; DB i3; Length 191;
Best Local Similarity 90.8%; Pred. No. 2.00e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKNTNRRPMDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKNTNRRPMDVKFPGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
|||||

Db 61 RRQIPKARSEGRSWAQPYPMPLYGNECGWAGWLLSPRGRSPWGP 109
|||||
Qy 61 RRQIPKARSEGRSWAQPYPMPLYGNECGWAGWLLSPRGRSPWGP 109
|||||

RESULT 3
ID PCT-US95-10398-158 STANDARD; PRT; 191 AA.
XX AC xxxxxx
XX 01-JAN-1900
XX
DE Sequence 158, Application PC/TUS9510398.
XX
CC Sequence 158, Application PC/TUS9510398

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ai

5

CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 158:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: SW1
CC SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRPQDVKFGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60
Qy 1 MSTNPKQRTKRNTRRPQDVKFGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60

Db 61 RQPIPKARPERGTWAPGYPFPLYNGECGWACHLLSPRGRSWGP 109

Qy 61 RQPIPKARPERGTWAPGYPFPLYNGECGWACHLLSPRGRSWGP 109

RESULT 4

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ai

6

ID PCT-US95-10398-156 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 156, Application PC/TUS9510398.
XX
CC Sequence 156, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 156:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: US11
CC SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRPQDVKFGGQIVGGVYLLPRRGLGVRATKTSERSQPRG 60

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

7

```
Qy 1 MSTNPKQKTRNTNRPRXXXXXPGGQIVGGVYLLPRGPRGXGVRATRTKTSERSQPRG 60
Db 61 RROPIKARPEGRTPAOPGYPWPPLYGNEGCGMACWLLSPRGSRPSWGP 109
Qy 61 RROPIKAXRGEGRSQAOPGYPWPPLYGNEGCGMACWLLSPRGSRPNWGP 109

RESULT 5
ID PCT-US95-10398-157 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 157, Application PC/TUS9510398.
XX
CC Sequence 157, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
```

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

8

```
CC ORGANISM: hom sapiens
CC INDIVIDUAL ISOLATE: S14
SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTRNTNRPRQDVKFPFGGQIVGGVYLLPRGPRGLGVRATRTKTSERSQPRG 60
Qy 1 MSTNPKQKTRNTNRPRXXXXXPGGQIVGGVYLLPRGPRGXGVRATRTKTSERSQPRG 60
Db 61 RROPIKARPEGRTPAOPGYPWPPLYGNEGCGMACWLLSPRGSRPSWGP 109
Qy 61 RROPIKAXRGEGRSQAOPGYPWPPLYGNEGCGMACWLLSPRGSRPNWGP 109

RESULT 6
ID PCT-US95-10398-196 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 196, Application PC/TUS9510398.
XX
CC Sequence 196, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
```

```

CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 196:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: 27
SQ SEQUENCE 191 AA; 20788 MW; 180939 CN;

Query Match 99.2%; Score 775; DB 13; Length 191;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQKTKNTNRMDVKEPGGGQIVGGVYLLPRGPRGLGVTRTKTTSRSQPRG 60
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 |||
Qy 1 MSTNPKPQKTKNTNRPEXXXXPFGGGQIVGGVYLLPRGPRGXGVTRTKTTSRSQPRG 60
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 |||
Db 61 RRQIPKARRSEGRSQAQGYPMPLYNGEGCGGACWLLSPRGSRPSWCP 109
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 |||
Qy 61 RRQIPKAXRSEGRSQAQGYPMPLYNGEGCGGACWLLSPRGSRPNWCP 109
||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 ||||||| 1 |||

RESULT 7
ID PCT-US95-10398-159 STANDARD; PRT; 191 AA.
XX AC xxxxxx
DT 01-JAN-1900
XX DE Sequence 159, Application PC/TUS9510398.
XX SE Sequence 159, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC PRIOR APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993

```

CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER: 08/290/665	
CC	FILING DATE: 15 AUGUST 1994	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: RICHARD W. BORK	
CC	REGISTRATION NUMBER: 36,459	
CC	REFERENCE/DOCKET NUMBER: 2026-4116	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (212) 758-4800	
CC	TELEFAX: (212) 751-6849	
CC	TELEX: 421792	
CC	INFORMATION FOR SEQ ID NO: 159:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 191 amino acids	
CC	TYPE: amino acid	
CC	STRANDEDNESS: unknown	
CC	TOPOLOGY: unknown	
CC	ORIGINAL SOURCE:	
CC	ORGANISM: hom sapiens	
CC	INDIVIDUAL ISOLATE: S18	
CC	SEQUENCE 191 AA; 20766 MW; 182089 CN;	
Qy	Query Match 99.2%; Score 775; DB 13; Length 191;	
Qy	Best Local Similarity 89.9%; Pred. No. 3.03e-59;	
Qy	Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;	
Db	1 MSTNPKQPKTKRNTNRRPDVKEPFGGQIVGCVYLLPRGPRLCVTRATKTSERSQPRG 60	
Qy	1 MSTNPKQPKTKRNTNRRPDVKEPFGGQIVGCVYLLPRGPRLCVTRATKTSERSQPRG 60	
Db	61 RRQIPKARRPEGRTWAQPGTPWPLVYNGCGGAGWLLSPGSRPSWGP 109	
Qy	61 RRQIPKARRPEGRTWAQPGTPWPLVYNGCGGAGWLLSPGSRPSWGP 109	
RESULT 8		
ID	PCT-US95-10398-160 STANDARD; PRY; 191 AA.	
XX	xxxxxx	
AC	R.H.	
XX	01-JAN-1900	
DT		
XX		
DE	Sequence 160, Application PC/TUS9510398.	
XX		
CC	Sequence 160, Application PC/TUS9510398	
CC	GENERAL INFORMATION:	
CC	APPLICANT: BUKH, J., MILLER, R.H. AND	
CC	APPLICANT: PURCELL, R.H.	
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED	
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND	
CC	TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS	
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE	
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES	
CC	NUMBER OF SEQUENCES: 263	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: MORGAN & FINNEGAN	
CC	STREET: 345 PARK AVENUE	
CC	CITY: NEW YORK	
CC	STATE: NEW YORK	
CC	COUNTRY: USA	
CC	ZIP: 10154	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: FLOPPY DISK	
CC	COMPUTER: IBM PC COMPATIBLE	

21

CC	STREET: ONE ABBOTT PARK ROAD
CC	CITY: ABBOTT PARK-
CC	STATE: IL
CC	COUNTRY: USA
CC	ZIP: 60064-3500
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: Floppy disk
CC	COMPUTER: IBM PC compatible
CC	OPERATING SYSTEM: PC-DOS/MS-D
CC	SOFTWARE: PatentIn Release #1
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: PCT/US94/
CC	FILING DATE:
CC	CLASSIFICATION:
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: POBEMSKI, PRISCILLA E.
CC	REGISTRATION NUMBER: 33,207
CC	REFERENCE/DOCKET NUMBER: 5521
CC	TELECOMMUNICATION INFORMATION:
CC	TELEPHONE: 708-937-6365
CC	TELEFAX: 708-938-2623
CC	INFORMATION FOR SEQ ID NO: 13:
CC	SEQUENCE CHARACTERISTICS:
CC	LENGTH: 967 amino acids
CC	TYPE: amino acid
CC	STRANDEDNESS: single
CC	TOPOLOGY: linear
CC	MOLECULE TYPE: protein
CC	SEQUENCE 967 AA; 106397 MW; 501417

[illegible]

RECORD ID	STANDARD	PRT
967 AA	PCT-US95-01087-13	

XX	01-JAN-1900
XX	
XX	
DE	Sequence 13, Application PC/TUS9501087.
XX	
CC	Sequence 13, Application PC/TUS9501087
CC	GENERAL INFORMATION:
CC	APPLICANT: WATANABE, SHINICHI
CC	APPLICANT: YAMAGUCHI, JULIE
CC	APPLICANT: DESAI, SURESH M.
CC	APPLICANT: DEVARE, SUSHIL G.
CC	TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC	TITLE OF INVENTION: ENVELOPE GENES
CC	NUMBER OF SEQUENCES: 22
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC	STREET: ONE ABBOTT PARK ROAD

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

13

CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/01087
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 967 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 967 AA; 106397 MW; 5014170 CN;

Query Match 99.2%; Score 775; DB 13; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTRKNTNRNRPQVKGQIVGGVYLLPRRCPLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTRKNTNRNRPQVKGQIVGGVYLLPRRCPLGVTRATKTSERSQPRG 60
|||||

Db 61 RRQPIKARPCGRTWAQPCYPMPLYGNEGCGWAGWLLSPRGSRPWSGP 109
|||||

Qy 61 RRQPIKARPCGRTWAQPCYPMPLYGNEGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 11
ID US-08-188-281B-13 STANDARD; PRT; 967 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 13, Application US/08188281B.
XX
CC Sequence 13, Application US/08188281B
CC Patent No. 5610009
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

14

CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,281B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 13:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 967 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 967 AA; 106397 MW; 5014170 CN;

Query Match 99.2%; Score 775; DB 7; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQKTRKNTNRNRPQVKGQIVGGVYLLPRRCPLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTRKNTNRNRPQVKGQIVGGVYLLPRRCPLGVTRATKTSERSQPRG 60
|||||

Db 61 RRQPIKARPCGRTWAQPCYPMPLYGNEGCGWAGWLLSPRGSRPWSGP 109
|||||

Qy 61 RRQPIKARPCGRTWAQPCYPMPLYGNEGCGWAGWLLSPRGSRPWSGP 109
|||||

RESULT 12
ID PCT-US95-01087-12 STANDARD; PRT; 1648 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 12, Application PC/TUS9501087.
XX
CC Sequence 12, Application PC/TUS9501087
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK

```

CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/01087
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

```

Query Match	99.2%	Score 775	DB 13	Length 1648
Best Local Similarity	89.9%	Pred. No. 3.03e-59		
Matches	98	Conservative 2	Mismatches 9	Indels 0
				Gaps 0
Db	1	MS1NPKPQKTKNTNRRPDVKFPGCGGIVGGVYLLPRRGRLGVYRTRKTSERSQPRG	60	
Qy	1	MS1NPKPQKTKNTNRRPDVKFPGCGGIVGGVYLLPRRGRLGVYRTRKTSERSQPRG	60	
Db	61	RROP1PKARRPGRWQPGYPWPVLYNCGCGGAGCWLSPRSRPSWGP	109	
Qy	61	RROP1PKAXRGRWQPGYPWPVLYNCGCGGAGCWLSPRSRPSWGP	109	

RESULT	13
ID	PCT-US94-07280-12
XX	STANDARD;
AC	XXXXXX
DT	01-JAN-1900
XX	
DE	Sequence 12, Application PC/TUS9407280.
XX	
CC	Sequence 12, Application PC/TUS9407280
CC	GENERAL INFORMATION:
CC	APPLICANT: WATANABE, SHINICHI
CC	APPLICANT: YAMAGUCHI, JULIE
CC	APPLICANT: DESAI, SURESH M.
CC	APPLICANT: DEVARE, SUSHIL G.
CC	TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS
CC	TITLE OF INVENTION: ENVELOPE GENES
CC	NUMBER OF SEQUENCES: 22
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSES: ABBOTT LABORATORIES D377/AP6D
CC	STREET: ONE ABBOTT PARK ROAD
CC	CITY: ABBOTT PARK
CC	STATE: IL

CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/07280
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC
SQ SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

	Query Match	99.2%	Score 775;	DB 12;	Length 1648;
	Best Local Similarity	89.9%;	Pred. No. 3.03e-59;		
	Matches	98;	Conservative	9;	Indels 0; Gaps 0;
Db	1	MSNTPEQKTKNTNRRPDVKPE	GGGQIVGGVYLLPRGRLCVTRTKT	SQPRG 60	
Qy	1	MSNTPEQKTKNTNRRPDVXXXXP	GGGQIVGGVYLLPRGRLCVTRTKT	SQPRG 60	
Db	61	RROP1PKARRPEGTHAQP	GYPWPLYGNEGCGCAWLL	SPGSRP	SWCP 109
Ov	61	RROP1PKARRPEGTHAQP	GYPWPLYGNEGCGCAWLL	SPGSRP	RMCP 109

[illegible]

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

17

CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/188,281B
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 12:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1648 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 1648 AA; 178869 MW; 14581021 CN;

Query Match 99.2%; Score 775; DB 7; Length 1648;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGCVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGCVTRATKTSERSQPRG 60
|||||
Db 61 RRQIPKARPRGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPSWGP 109
|||||
Qy 61 RRQIPKARPRGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPSWGP 109

RESULT 15
ID PCT-US94-07280-1 STANDARD; PRT; 3011 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 1, Application PC/TUS9407280.
XX
CC Sequence 1, Application PC/TUS9407280
CC GENERAL INFORMATION:
CC APPLICANT: WATANABE, SHINICHI
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC NUMBER OF SEQUENCES: 22
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA

Feb 23 11:03

/home/fuller/feb1098/US-08-836-075A-2.ra

18

CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US94/07280
CC FILING DATE:
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5521.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-938-2623
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 3011 AA; 327108 MW; 47577590 CN;

Query Match 99.2%; Score 775; DB 12; Length 3011;
Best Local Similarity 89.9%; Pred. No. 3.03e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGCVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRPPQDKFPGGGQIVGGVYLLPRGPRGCVTRATKTSERSQPRG 60
|||||
Db 61 RRQIPKARPRGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPSWGP 109
|||||
Qy 61 RRQIPKARPRGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPSWGP 109

Search completed: Mon Feb 23 11:06:01 1998
Job time : 12 secs.

[illegible]

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by Intelligenetics, Inc.

protein - protein database search, using Smith-Waterman algorithm

```
Run on: Mon Feb 23 11:03:09 1998; MasPar time 11.95 Seconds
226.191 Million cell updates/sec
```

Tabular output not generated.

```

>US-08-836-075A-2
Description: (1-109) from US08836075A.pap
Perfect Score: 781
Sequence: 1 MSTNPKPQKTKRNTNRRPX.....
Scoring table: PAM 150
Gap 11

```

Searched: 243240 seqs, 24799054 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database:

a-pending

1: PCT91 2: PCT92 3: PCT93 4: PCT94 5: PCT95 6: PCT96 7: PCT97
8: PCT98 9: U76 10: U77A 11: U77B 12: U78A 13: U78B 14: U79A
15: U79B 16: U800 17: U801 18: U802 19: U803 20: U804 21: U805
22: U806 23: U807 24: U808 25: U809 26: U810 27: U811 28: U812
29: U813 30: U814 31: U815 32: U816 33: U817 34: U818 35: U819
36: U820 37: U821 38: U822 39: U823 40: U824 41: U825 42: U826
43: U827 44: U828 45: U829 46: U830 47: U831 48: U832 49: U833
50: U834 51: U835 52: U836 53: U837 54: U838 55: U839 56: U840
57: U841 58: U842 59: U843 60: U844 61: U845 62: U846 63: U847
64: U848 65: U849 66: U85 67: U856 68: U857 69: U88 70: U89
71: U60 72: NF04 73: NF06 74: NF08
75: NF09

Statistics: Mean 27.872: Variance 128.076: scale 0.218

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query Match	Length	DB	ID	Description	Pred. No.
1	781	100.0	109	74	US-08-836-Sequence 2, Applicatio	5.66e-60		
2	777	99.5	117	74	US-08-836-Sequence 28, Applicati	1.32e-59		
3	777	99.5	191	45	US-08-290-Sequence 195, Applicat	1.32e-59		
4	777	99.5	191	45	US-08-290-Sequence 192, Applicat	1.32e-59		

CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836,075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:004
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 109 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 109 AA; 12273 MW; 64913 CN;

Query Match 100.0%; Score 781; DB 74; Length 109;
Best Local Similarity 100.0%; Pred. No. 5.66e-60;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPXXXXPGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPXXXXPGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||

Db 61 RRQPIKAXRXEGRSWAQPGYPWPPLYNCGCGMAXWLLSPRGRNWP 109
|||||
Qy 61 RRQPIKAXRXEGRSWAQPGYPWPPLYNCGCGMAXWLLSPRGRNWP 109
|||||

RESULT 2
ID US-08-836-075A-28 STANDARD; PRT; 117 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 28, Application US/08836075A.
XX
CC Sequence 28, Application US/08836075A
CC GENERAL INFORMATION:
CC APPLICANT: MAERTENS, GEERT
CC APPLICANT: STUVER, LIEVEN
CC TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
CC TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DI
AGNOSTIC

CC TITLE OF INVENTION: AGENTS
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836,075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:004
CC INFORMATION FOR SEQ ID NO: 28:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 117 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: peptide
SQ SEQUENCE 117 AA; 13380 MW; 67911 CN;

Query Match 99.5%; Score 777; DB 74; Length 117;
Best Local Similarity 91.7%; Pred. No. 1.32e-59;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPMDVKFPGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPXXXXPGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
|||||

Db 61 RRQPIKAXRXEGRSWAQPGYPWPPLYNCGCGMAXWLLSPRGRNWP 109
|||||
Qy 61 RRQPIKAXRXEGRSWAQPGYPWPPLYNCGCGMAXWLLSPRGRNWP 109
|||||

RESULT 3
ID US-08-290-665A-195 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 195, Application US/08290665A.
XX
CC Sequence 195, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUXH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 195:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: Z6
SQ SEQUENCE 191 AA; 20774 MW; 184617 CN;

Query Match 99.5%; Score 777; DB 45; Length 191;
Best Local Similarity 90.8%; Pred. No. 1.32e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKPQRTKNTNRRPMDVKFCGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Qy 1 MSTNPKPQRTKNTNRRPMDVKFCGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Db 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGSRPSWGP 109
Qy 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGSRPNWGP 109

RESULT 4
ID US-08-290-665A-192 STANDARD; PRT; 191 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX Sequence 192, Application US/08290665A.
XX
XX Sequence 192, Application US/08290665A
CC

CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 192:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: Z8
SQ SEQUENCE 191 AA; 20744 MW; 181593 CN;

Query Match 99.5%; Score 777; DB 45; Length 191;
Best Local Similarity 90.8%; Pred. No. 1.32e-59;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKPQRTKNTNRRPMDVKFCGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Qy 1 MSTNPKPQRTKNTNRRPMDVKFCGGQIVGGVYLLPRGPRLGVRATRTKTSERSQPRG 60
Db 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGSRPSWGP 109
Qy 61 RRQPIPKARRSEGRSWAQPYPWPLYGNEGGCGWAGLLSPRGSRPNWGP 109

RESULT 5
ID US-07-786-169A-6 STANDARD; PRT; 120 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
XX

DE Sequence 6, Application US/07786169A.

XX Sequence 6, Application US/07786169A

CC GENERAL INFORMATION:

CC APPLICANT: Krupka, Udo

CC APPLICANT: Stuber, Werner

CC APPLICANT: Gerken, Manfred

CC APPLICANT: Brust, Stefan

CC TITLE OF INVENTION: HCV-Specific Peptides, Agents Therefor

CC TITLE OF INVENTION: and the Use Thereof

CC NUMBER OF SEQUENCES: 44

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &

CC ADDRESSEE: Dunner

CC STREET: 1300 I Street, N.W.

CC CITY: Washington

CC STATE: D.C.

CC COUNTRY: USA

CC ZIP: 20005-3315

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/07786,169A

CC FILING DATE: 19911031

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Kirschner, Michael K.

CC REGISTRATION NUMBER: 34,851

CC REFERENCE/DOCKET NUMBER: 02481-1118-00000

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: 202-408-4000

CC TELEFAX: 202-408-4400

CC INFORMATION FOR SEQ ID NO: 6:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 120 amino acids

CC TYPE: AMINO ACID

CC TOPOLOGY: linear

CC MOLECULE TYPE: peptide

SQ SEQUENCE 120 AA; 13619 MW; 70513 CN;

Query Match 99.2%; Score 775; DB 11; Length 120;

Best Local Similarity 89.9%; Pred. No. 2.01e-59;

Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQRTKRNTRRPQDKVFGGGQIVGGVYLLPRGPRGLGVTRTKTSERSQPRG 60

QY 1 MSTNPKPQRTKRNTRRPQDKVFGGGQIVGGVYLLPRGPRGXGVRATKTSERSQPRG 60

Db 61 RRQPIPKARPEGRTPAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPWSGP 109

QY 61 RRQPIPKAXRKGSRWAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPNWGP 109

RESULT 6

ID US-08-290-665A-158 STANDARD; PRT; 191 AA.

XX xxxxxx

XX 01-JAN-1900

XX DE

DE Sequence 158, Application US/08290665A.

XX Sequence 158, Application US/08290665A

CC GENERAL INFORMATION:

CC APPLICANT: BURKH, J., MILLER, R.H. AND

CC APPLICANT: PURCELL, R.H.

CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED

CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND

CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS

CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE

CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES

CC NUMBER OF SEQUENCES: 263

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: MORGAN & FINNEGAN

CC STREET: 345 PARK AVENUE

CC CITY: NEW YORK

CC STATE: NEW YORK

CC COUNTRY: USA

CC ZIP: 10154

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: FLOPPY DISK

CC COMPUTER: IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: WORDPERFECT 5.1

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/290,665A

CC FILING DATE: 15-AUG-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: RICHARD W. BORK

CC REGISTRATION NUMBER: 36,459

CC REFERENCE/DOCKET NUMBER: 2026-4116

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (212) 758-4800

CC TELEFAX: (212) 751-6849

CC TELEX: 421792

CC INFORMATION FOR SEQ ID NO: 158:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 191 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: unknown

CC TOPOLOGY: unknown

CC ORIGINAL SOURCE:

CC ORGANISM: homosapiens

CC INDIVIDUAL ISOLATE: SW1

SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;

Best Local Similarity 89.9%; Pred. No. 2.01e-59;

Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKPQRTKRNTRRPQDKVFGGGQIVGGVYLLPRGPRGLGVTRTKTSERSQPRG 60

QY 1 MSTNPKPQRTKRNTRRPQDKVFGGGQIVGGVYLLPRGPRGXGVRATKTSERSQPRG 60

Db 61 RRQPIPKARPEGRTPAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPWSGP 109

QY 61 RRQPIPKAXRKGSRWAQPGYPWPPLYGNEGCGWAGWLLSPRGSRPNWGP 109

RESULT 7

ID US-08-290-665A-196 STANDARD; PRT; 191 AA.

XX xxxxxx

XX AC

XX DE

DT 01-JAN-1900
XX
DE Sequence 196, Application US/08290665A.
XX
CC Sequence 196, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 196:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: 27
CC SEQUENCE 191 AA; 20788 MW; 180939 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Dd 1 MSTNPKPQRTKRNTRRRPDVKFPGGGQIVGGVYLLPRGPRGLGVRTRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKPQRTKRNTRRRPDVKFPGGGQIVGGVYLLPRGPRGLGVRTRTKTSERSQPRG 60
|||||

Dd 61 RROIPKARRSEGRSWAQPYPMPLYGNEGCGWAGHLLSPRGSRP SWGP 109
|||||

Qy 61 RRQIPKAXRXEGRSWAQPYPMPLYGNEGCGWAGHLLSPRGSRP SWGP 109
|||||

RESULT 8
ID US-08-380-160-3 STANDARD; PRT; 191 AA.

XX
AC xxxxxx
XX 01-JAN-1900
XX
DE Sequence 3, Application US/08380160.
XX
CC Sequence 3, Application US/08380160
CC GENERAL INFORMATION:
CC APPLICANT: DALBON, Pascal
CC APPLICANT: JOLIVET, Michel
CC TITLE OF INVENTION: SYNTHETIC POLYPEPTIDES BELONGING TO THE
CC TITLE OF INVENTION: HEPATITIS C VIRUS (HCV) AND WHICH CAN BE USED ESP
CC
CC Specially
CC TITLE OF INVENTION: FOR DETECTING THE LATTER
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: OLIFF & BERRIDGE
CC STREET: P.O. Box 19928
CC CITY: Alexandria
CC STATE: VA
CC COUNTRY: USA
CC ZIP: 22320
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/380,160
CC FILING DATE:
CC CLASSIFICATION: 530
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/057,471
CC FILING DATE: 06-MAY-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Berridge, William P.
CC REGISTRATION NUMBER: 30,024
CC REFERENCE/DOCKET NUMBER: WPB 28682
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (703)836-6400
CC TELEFAX: (703)836-2787
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Human Hepatitis C Virus
CC SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 54; Length 191;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Dd 1 MSTNPKPQRTKRNTRRRPDVKFPGGGQIVGGVYLLPRGPRGLGVRTRTKTSERSQPRG 60
|||||
Qy 1 MSTNPKPQRTKRNTRRRPDVKFPGGGQIVGGVYLLPRGPRGLGVRTRTKTSERSQPRG 60
|||||

* Best Local Similarity 89.94; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RROPIKARRECGRTWAQPCYPMPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RROPIKARRECGRTWAQPCYPMPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||

RESULT 11
ID US-08-290-665A-157 STANDARD; PRT; 191 AA.

XX
AC xxxxxx
DT 01-JAN-1900
XX
DE Sequence 157, Application US/08290665A.
CC
CC Sequence 157, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 157:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: unknown
CC TOPOLOGY: unknown
CC ORIGINAL SOURCE:
CC ORGANISM: homoplans
CC INDIVIDUAL ISOLATE: S14

SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.28; Score 775; DB 45; Length 191;
Best Local Similarity 89.94; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 MSTNPKQRTKRNTRRRPQDKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQRTKRNTRRRPQDKFPGGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||

Db 61 RROPIKARRECGRTWAQPCYPMPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RROPIKARRECGRTWAQPCYPMPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||

RESULT 12
ID US-08-290-665A-156 STANDARD; PRT; 191 AA.

XX
AC xxxxxx
DT 01-JAN-1900
XX
DE Sequence 156, Application US/08290665A.
CC
CC Sequence 156, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290,665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 156:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 191 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: unknown

Feb 23 11:02

/home/fuller/feb1098/US-08-836-075A-2.rap

15

CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: US11
SQ SEQUENCE 191 AA; 20766 MW; 182089 CN;

Query Match 99.2%; Score 775; DB 45; Length 191;
Best Local Similarity 89.9%; Pred. No. 2.0le-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Dbb 1 MSTNPKQKTKRNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRRPXXXXXFGGQIVGGVYLLPRGPRGXGVTRATKTSERSQPRG 60
|||||

Dbb 61 RRQPIPKARRPEGRTHAQPQGYWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQPIPKAXXEGRSWAQGYWPLYGNECGGAXWLLSPRGSRPNWGP 109
|||||

RESULT 13
ID US-08-224-973A-23 STANDARD; PRT; 200 AA.
XX
AC xxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 23, Application US/08224973A.
XX
CC Sequence 23, Application US/08224973A
CC GENERAL INFORMATION:
CC APPLICANT: Berzofsky, Jay A.
CC APPLICANT: Feinstein, Stephen
CC APPLICANT: Shirai, Mutsunori
CC TITLE OF INVENTION: Hepatitis C Virus Core Peptide for
CC TITLE OF INVENTION: Stimulation of Cytotoxic T Lymphocytes and Diagno
sis of
CC TITLE OF INVENTION: HCV Exposure
CC NUMBER OF SEQUENCES: 26
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Birch, Stewart, Kolasch & Birch, LLP
CC STREET: P.O. Box 747
CC CITY: Falls Church
CC STATE: Virginia
CC COUNTRY: USA
CC ZIP: 22040-0747
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/224,973A
CC FILING DATE: 08-APR-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Svensson, Leonard R.
CC REGISTRATION NUMBER: 30330
CC REFERENCE/DOCKET NUMBER: 1173-456P
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 703-205-8000
CC TELEFAX: 703-205-8050
CC INFORMATION FOR SEQ ID NO: 23:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 200 amino acids
CC TYPE: amino acid

Feb 23 11:02 /home/fuller/feb1098/US-08-836-075A-2.rap

16

CC STRANDEDNESS: not relevant
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC HYPOTHETICAL: NO
CC FRAGMENT TYPE: N-terminal
CC ORIGINAL SOURCE:
CC ORGANISM: Hepatitis C Virus
CC STRAIN: NYBC
CC FEATURE:
CC NAME/KEY: Protein
CC LOCATION: 1..200
CC OTHER INFORMATION: /note= "HCV core protein, NYBC
CC OTHER INFORMATION: isolate, see Fig. 5"
SQ SEQUENCE 200 AA; 21771 MW; 201865 CN;

Query Match 99.2%; Score 775; DB 38; Length 200;
Best Local Similarity 89.9%; Pred. No. 2.0le-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Dbb 1 MSTNPKQKTKRNTNRRPQDVKFGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
|||||
Qy 1 MSTNPKQKTKRNTNRRPXXXXXFGGQIVGGVYLLPRGPRGXGVTRATKTSERSQPRG 60
|||||

Dbb 61 RRQPIPKARRPEGRTHAQPQGYWPLYGNECGGAGWLLSPRGSRPWSGP 109
|||||
Qy 61 RRQPIPKAXXEGRSWAQGYWPLYGNECGGAXWLLSPRGSRPNWGP 109
|||||

RESULT 14
ID US-08-710-637-1 STANDARD; PRT; 3011 AA.
XX
AC xxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 1, Application US/08710637.
XX
CC Sequence 1, Application US/08710637
CC GENERAL INFORMATION:
CC APPLICANT: CASEY, JAMES M.
CC APPLICANT: BODE, SUZANNE L.
CC APPLICANT: ZECK, BILLY J.
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: FRAIL, DONALD E.
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC TITLE OF INVENTION: PROTEINS
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP6D
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/710,637
CC FILING DATE:

Feb 23 11:02

/home/fuller/feb1098/US-08-836-075A-2.rap

17

CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/144,099
CC FILING DATE:
CC APPLICATION NUMBER: US 07/830,024
CC FILING DATE: 01-JAN-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5131.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-937-9556
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 3011 AA; 327108 MW; 47577590 CN;

Query Match 99.2%; Score 775; DB 58; Length 3011;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQRTKNTNRRPDVKFGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
QY 1 MSTNPKQRTKNTNRRPDVXXXXXPGGGQIVGGVYLLPRGPRGXGVRATKTSERSQPRG 60
Db 61 RRQPIPKARPEGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPWSGP 109
QY 61 RRQPIPKAXRGEGRSWAQPGYWPPLYGNECGGAGWLLSPRGSRPNWGP 109

RESULT 15
ID US-08-710-637-2 STANDARD; PRT; 3011 AA.
XX
AC xxxxxx
XX
DT 01-JAN-1900
XX
DE Sequence 2, Application US/08710637.
XX
CC Sequence 2, Application US/08710637
CC GENERAL INFORMATION:
CC APPLICANT: CASEY, JAMES M.
CC APPLICANT: BODE, SUZANNE L.
CC APPLICANT: ZECK, BILLY J.
CC APPLICANT: YAMAGUCHI, JULIE
CC APPLICANT: FRAIL, DONALD E.
CC APPLICANT: DESAI, SURESH M.
CC APPLICANT: DEVARE, SUSHIL G.
CC TITLE OF INVENTION: MAMMALIAN EXPRESSION SYSTEMS FOR HCV
CC TITLE OF INVENTION: PROTEINS
CC NUMBER OF SEQUENCES: 12
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ABBOTT LABORATORIES D377/AP60
CC STREET: ONE ABBOTT PARK ROAD
CC CITY: ABBOTT PARK
CC STATE: IL
CC COUNTRY: USA
CC ZIP: 60064-3500
CC COMPUTER READABLE FORM:

Feb 23 11:02

/home/fuller/feb1098/US-08-836-075A-2.rap

18

CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/710,637
CC FILING DATE:
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/144,099
CC FILING DATE:
CC APPLICATION NUMBER: US 07/830,024
CC FILING DATE: 01-JAN-1992
CC ATTORNEY/AGENT INFORMATION:
CC NAME: POREMBSKI, PRISCILLA E.
CC REGISTRATION NUMBER: 33,207
CC REFERENCE/DOCKET NUMBER: 5131.US.01
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 708-937-6365
CC TELEFAX: 708-937-9556
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 3011 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
SQ SEQUENCE 3011 AA; 327194 MW; 47602524 CN;

Query Match 99.2%; Score 775; DB 68; Length 3011;
Best Local Similarity 89.9%; Pred. No. 2.01e-59;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Db 1 MSTNPKQRTKNTNRRPDVKFGGQIVGGVYLLPRGPRGLGVTRATKTSERSQPRG 60
QY 1 MSTNPKQRTKNTNRRPDVXXXXXPGGGQIVGGVYLLPRGPRGXGVRATKTSERSQPRG 60
Db 61 RRQPIPKARPEGRTWAQPGYWPPLYGNECGGAGWLLSPRGSRPWSGP 109
QY 61 RRQPIPKAXRGEGRSWAQPGYWPPLYGNECGGAGWLLSPRGSRPNWGP 109

Search completed: Mon Feb 23 11:05:31 1998
Job time : 142 secs.

Db 61 rriqpikarpegrtwagpypwlyngcwgagwllaprgsrpswp 109
QY 61 RRQIPKAXRXXRGSWAQPGYPMPLYGNECGGMAXLLSPRGSRPNWGP 109

RESULT 3
ID POLG HCVJA STANDARD; PRT; 3010 AA.
AC P2662;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA
DE POLYMERASE (EC 2.7.7.48) (NS5)).
OS HEPATITIS C VIRUS (ISOLATE JAPANESE) (HCV).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC HCV GROUP VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 9108550.
RA KATO N., HIJIKATA M., OOTSUYAMA Y., NAKAGAWA M., OHKOSHI S.,
RA SUGIMURA T., SHIMOTOHNO K.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:9524-9528 (1990).
RN [2]
RP SEQUENCE ANALYSIS.
RX MEDLINE; 91192160.
RA KATO N., HIJIKATA M., NAKAGAWA M., OOTSUYAMA Y., MURAIKO S.,
RA OHKOSHI S., SHIMOTOHNO K.;
RL FEBS LETT. 280:325-328 (1991).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
DR EMBL; D90208; G221611; -.
DR PIR; A39253; GNMVJ.
KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
KW NONSTRUCTURAL PROTEIN.
FT INIT MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 2 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT CHAIN 730 1006 NON-STRUCTURAL PROTEIN NS2 (POTENTIAL).
FT CHAIN 1007 1615 HELICASE (NS3) (POTENTIAL).
FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (NS5)
(POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT NP BIND 1230 1237 ATP (POTENTIAL).
FT SITE 1316 1319 DECH BOX.
FT CARBOHYD 196 196 POTENTIAL.
FT CARBOHYD 209 209 POTENTIAL.
FT CARBOHYD 234 234 POTENTIAL.
FT CARBOHYD 250 250 POTENTIAL.
FT CARBOHYD 305 305 POTENTIAL.
FT CARBOHYD 417 417 POTENTIAL.

FT CARBOHYD 423 423 POTENTIAL.
FT CARBOHYD 430 430 POTENTIAL.
FT CARBOHYD 448 448 POTENTIAL.
FT CARBOHYD 532 532 POTENTIAL.
FT CARBOHYD 556 556 POTENTIAL.
FT CARBOHYD 576 576 POTENTIAL.
FT CARBOHYD 623 623 POTENTIAL.
FT CARBOHYD 645 645 POTENTIAL.
FT CARBOHYD 2041 2041 POTENTIAL.
FT CARBOHYD 2077 2077 POTENTIAL.
FT CARBOHYD 2240 2240 POTENTIAL.
FT CARBOHYD 2788 2788 POTENTIAL.
SQ SEQUENCE 3010 AA; 327017 MW; C2F2E630 CRC32;

Query Match 96.5%; Score 754; DB 7; Length 3010;
Best Local Similarity 89.0%; Pred. No. 6.16e-123;
Matches 97; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 1 mstnkpqkttkrntnrrpdkvkgpgggglvggyvllprgprlgvratktksersqrg 60
QY 1 MSTNPKQKTKRNTNRPRYXXXXXPGGGQIVGGVYLLPRRGPRXGVRATKRTKTSERSQPRG 60
Db 61 rriqpikarpegrtwagpypwlyngcwgagwllaprgsrpswp 109
QY 61 RRQIPKAXRXXRGSWAQPGYPMPLYGNECGGMAXLLSPRGSRPNWGP 109

RESULT 4
ID POLG HCVJ2 STANDARD; PRT; 513 AA.
AC P27959;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
DE PROTEIN NS1) (FRAGMENT).
OS HEPATITIS C VIRUS (ISOLATE HC-J2) (HCV).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
OC HCV GROUP VIRUSES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92230232.
RA OKAMOTO H., KURAI K., OKADA S.I., YAMAMOTO K., LIZUKA H.,
RA TANAKA T., FUKUDA S., TSUDA F., MISHIRO S.;
RL VIROLOGY 188:331-341 (1992).
CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
CC PROTEIN C AND RNA.
DR EMBL; D10074; G221589; -.
KW POLYPROTEIN; GLYCOPROTEIN; COAT PROTEIN; ENVELOPE PROTEIN;
KW TRANSMEMBRANE; NONSTRUCTURAL PROTEIN.
FT INIT MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
FT CELLULAR AMINOPEPTIDASE.
FT CHAIN 2 115 CAPSID PROTEIN C (POTENTIAL).
FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
FT CHAIN 384 >513 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
FT TRANSMEM 347 369 POTENTIAL.
FT CARBOHYD 196 196 POTENTIAL.
FT CARBOHYD 209 209 POTENTIAL.

Db 61 rrrqplkrrpawagpwpvlyneglgwaqwlleprgrspswcp 109
 QY 61 RRQIPKAXRSGRSGWAQGYWPLVYNEGCGWAXWLLSPRGSRPNWCP 109

RESULT 12
 ID POLG HCVJ6 STANDARD; PRT; 3033 AA.
 AC P26660;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
 DE PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA
 DE POLYMERASE (EC 2.7.7.48) (NS5)).
 OS HEPATITIS C VIRUS (ISOLATE HC-J6) (HCV).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
 OC HCV GROUP VIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 9204440.
 RA OKAMOTO H., OKADA S.-I., SUGIYAMA Y., KURAI K., LIZUKA H.,
 RA MACHIDA A., MIYAKAWA Y., MAYUMI M.;
 RL J. GEN. VIROL. 72:2697-2704(1991).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 DR EMBL; D00944; G221651; -.
 DR PIR; J01303; J01303.
 DR HSP; P01607; IFVC.
 KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
 KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
 KW NONSTRUCTURAL PROTEIN.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 2 115 CAPSID PROTEIN C (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 733 NONSTRUCTURAL PROTEIN NS1 (POTENTIAL).
 FT CHAIN 734 1010 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1011 1619 HELICASE (NS3) (POTENTIAL).
 FT CHAIN 1620 1866 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1867 2017 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2018 3033 RNA-DIRECTED RNA POLYMERASE (NS5)
 (POTENTIAL).
 FT TRANSMEM 347 369
 FT NP_BIND 1234 1241 ATP (POTENTIAL).
 FT SITE 1320 1323 DECH BOX.
 FT CARBOHYD 196 196 POTENTIAL.
 FT CARBOHYD 209 209 POTENTIAL.
 FT CARBOHYD 234 234 POTENTIAL.
 FT CARBOHYD 305 305 POTENTIAL.
 FT CARBOHYD 417 417 POTENTIAL.
 FT CARBOHYD 423 423 POTENTIAL.
 FT CARBOHYD 430 430 POTENTIAL.
 FT CARBOHYD 448 448 POTENTIAL.
 FT CARBOHYD 477 477 POTENTIAL.
 FT CARBOHYD 534 534 POTENTIAL.
 FT CARBOHYD 542 542 POTENTIAL.

FT CARBOHYD 558 558 POTENTIAL.
 FT CARBOHYD 578 578 POTENTIAL.
 FT CARBOHYD 627 627 POTENTIAL.
 FT CARBOHYD 649 649 POTENTIAL.
 FT CARBOHYD 1091 1091 POTENTIAL.
 FT CARBOHYD 2038 2038 POTENTIAL.
 FT CARBOHYD 2811 2811 POTENTIAL.
 SQ SEQUENCE 3033 AA; 329165 MW; 33355AC9 CRC32;

Query Match 93.2%; Score 728; DB 7; Length 3033;
 Best Local Similarity 85.3%; Pred. No. 8.70e-118;
 Matches 93; Conservative 3; Mismatches 13; Indels 0; Gaps 0;

Db 1 metnqpqrktkrntrrpqdvkfgggqivggvylrrprgrlgrvratrktsersqrg 60
 QY 1 MSTNPKQKTKRNTNRRPXXXXXPGGGQIVGGVYLLPRGPRXGVRATRKTSERSQPRG 60
 Db 61 rrrqplkrrtstgskwkpqywpvlyneglgwaqwlleprgrspswcp 109
 QY 61 RRQIPKAXRSGRSGWAQGYWPLVYNEGCGWAXWLLSPRGSRPNWCP 109

RESULT 13
 ID POLG HCVTW STANDARD; PRT; 3010 AA.
 AC P29846;
 DT 01-APR-1993 (REL. 25, CREATED)
 DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
 DT 01-FEB-1994 (REL. 28, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
 DE PROTEINS NS1, NS2, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED RNA
 DE POLYMERASE (EC 2.7.7.48) (NS5)).
 OS HEPATITIS C VIRUS (ISOLATE TAIWAN) (HCV).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE;
 OC HCV GROUP VIRUSES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 92230206.
 RA CHEN P.-J., LIN M.-H., TAI K.-F., LIU P.-C., LIN C.-J., CHEN D.-S.;
 RL VIROLOGY 188:102-113(1992).
 CC -1- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -1- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 DR EMBL; M484754; -; NOT ANNOTATED_CDS.
 DR PIR; A40244; GNWWTW.
 KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
 KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
 KW NONSTRUCTURAL PROTEIN.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 CELLULAR AMINOPEPTIDASE.
 FT CHAIN 2 115 CORE PROTEIN (POTENTIAL).
 FT CHAIN 116 191 MATRIX PROTEIN (POTENTIAL).
 FT CHAIN 192 383 MAJOR ENVELOPE PROTEIN E (POTENTIAL).
 FT CHAIN 384 729 NONSTRUCTURAL PROTEIN NS1/E2 (POTENTIAL).
 FT CHAIN 730 1006 NONSTRUCTURAL PROTEIN NS2 (POTENTIAL).
 FT CHAIN 1007 1615 HELICASE (NS3) (POTENTIAL).
 FT CHAIN 1616 1862 NONSTRUCTURAL PROTEIN NS4A (POTENTIAL).
 FT CHAIN 1863 2013 NONSTRUCTURAL PROTEIN NS4B (POTENTIAL).
 FT CHAIN 2014 3010 RNA-DIRECTED RNA POLYMERASE (NS5).
 FT TRANSMEM 347 369 POTENTIAL.

FT	NP	BIND	1230	1237	ATP (POTENTIAL).
FT	SITE		1316	1319	DECH BOX.
FT	CARBOHYD		196	196	POTENTIAL.
FT	CARBOHYD		209	209	POTENTIAL.
FT	CARBOHYD		233	233	POTENTIAL.
FT	CARBOHYD		234	234	POTENTIAL.
FT	CARBOHYD		250	250	POTENTIAL.
FT	CARBOHYD		305	305	POTENTIAL.
FT	CARBOHYD		417	417	POTENTIAL.
FT	CARBOHYD		423	423	POTENTIAL.
FT	CARBOHYD		430	430	POTENTIAL.
FT	CARBOHYD		448	448	POTENTIAL.
FT	CARBOHYD		532	532	POTENTIAL.
FT	CARBOHYD		540	540	POTENTIAL.
FT	CARBOHYD		556	556	POTENTIAL.
FT	CARBOHYD		576	576	POTENTIAL.
FT	CARBOHYD		623	623	POTENTIAL.
FT	CARBOHYD		645	645	POTENTIAL.
FT	CARBOHYD		2041	2041	POTENTIAL.
FT	CARBOHYD		2077	2077	POTENTIAL.
FT	CARBOHYD		2240	2240	POTENTIAL.
FT	CARBOHYD		2529	2529	POTENTIAL.
FT	CARBOHYD		2788	2788	POTENTIAL.
FT	SEQUENCE		3010 AA;	327047 MW;	47882A11 CRC32;

Query Match	92.8%	Score 725	DB 7	Length 3010
Best Local Similarity	86.2%	Pred. No. 3.41e-117	Indels 0	Gaps 0
Matches	94	Conservative	3	Mismatches 12
Db	1	metnqpkrtkntnrpqdvkfpggggagvgvylprqrglqvratrktwersqrg	60	
Qy	1	MTNPKPQRTKNTNRPRGXXXXPGGGGVLPRRGSRGXGVRATRKTSRQPRG	60	
Db	61	rcrpiptkarqpcgrawagpygpylyngedlgawglwqgylprgsrpnwqp	109	
Qy	61	RQRP LPKAXRXRGSRWAGPYGPMPLVNGEGCGXAWTLSPGSRGNWGP	109	

RESULT	14	ID	POLN RUBVT	STANDARD;	PRT;	2205 AA.
AC	P13889;	01-JAN-1990	(REL. 13, CREATED)			
DT	01-AUG-1992	(REL. 23, LAST SEQUENCE UPDATE)				
DT	01-AUG-1992	(REL. 23, LAST ANNOTATION UPDATE)				
DE	NONSTRUCTURAL POLYPROTEIN (CONTAINS: NONSTRUCTURAL PROTEINS NSP1 TO NSP4).					
DE						
OS	ROBELLA VIRUS (STRAIN THERIEN).					
OC	IRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; TOGAVIRIDAE;					
OC	RUBIVIRUSES.					
RN	[1]					
RP	SEQUENCE FROM N.A.					
RX	MEDLINE; 90281585.					
RA	DOMINGUEZ G., WANG C.Y., FREY T.K.;					
RL	VIROLOGY 177:225-258(1990).					
RN	[2]					
RP	SEQUENCE OF 1737-2205 FROM N.A.					
RX	MEDLINE; 88226020.					
RA	FREY T.K.; MARR L.D.;					
RL	GENE 62:85-99(1988).					
CC	-1- PTM: SPECIFIC ENZYMAIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS					
DR	EMBL; M15240; G333972; -.					
DR	PIR; A35320; MNVRN.					
FW	POLYPROTEIN; NONSTRUCTURAL PROTEIN.					
KT	CHAIN 1 ? NONSTRUCTURAL PROTEIN NSP1.					

FT	CHAIN	?	?		NONSTRUCTURAL PROTEIN NSP2.
FT	CHAIN	?	?		NONSTRUCTURAL PROTEIN NSP3.
FT	CHAIN	?	?	2205	NONSTRUCTURAL PROTEIN NSP4.
SQ	SEQUENCE	2205 AA;	240220 MW;	E2158BA CRC32;	
 Query Match 13.3%; Score 104; DB 7; Length 2205; Best Local Similarity 33.3%; Pred. No. 1.10e-02; Matches 15; Conservative 13; Mismatches 14; Indels 3; Gaps 3;					
Db	613 agardl-gafvrvvvppperwadgarawakffr-gcawaqtlhg 655	: :: ::: :	:: : :: :	:: : :: :	
Qy	56 SQRGRROP IPKAXRXEGRSWAQGY-PWPLYGNECGMAXWLLS 99	:	:: : :: :	:: : :: :	
 RESULT 15					
ID	CDP CANFA	STANDARD;	PRT;	975 AA.	
AC	P39881				
DT	01-FEB-1995 (REL. 31, CREATED)				
DT	01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)				
DT	01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)				
DE	CCAT DISPLACEMENT PROTEIN (HOMEBOX PROTEIN CLOX) (CLOX-1)				
DE	(FRAGMENT).				
GN	CUTL1 OR CLOX.				
OS	CANIS FAMILIARIS (DOG).				
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;				
OC	EUTHERIA; CARNIVORA.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=CARDIAC VENTRICULAR MUSCLE;				
RA	MEDLINE; 93161940.				
RX	ANDRES V., NADAL-GINARD B., MAHDAVI V.;				
RL	DEVELOPMENT 116:321-334 (1992).				
CC	-!- FUNCTION: DNA-BINDING REGULATORY PROTEIN, WHERE THE SMALLER				
CC	PROTEIN SPECIES PROBABLY REPRESENTS THE ACTIVE DNA-BINDING FORM.				
CC	MAY FUNCTION AS A DEVELOPMENTALLY REGULATED REPRESSOR OF TISSUE-				
CC	SPECIFIC GENE TRANSCRIPTION BY PREVENTING THE INTERACTION OF				
CC	TISSUE-SPECIFIC TRANSACTIVATORS WITH THEIR COGNATE TARGET				
CC	SEQUENCES. PROBABLY INVOLVED IN CELL-FATE SPECIFICATION IN DIVERSE				
CC	DIFFERENTIATION PROGRAMMES.				
CC	-!- DEVELOPMENTAL STAGE: DIFFERENTIALLY EXPRESSED DURING DEVELOPMENT.				
CC	SMALL PROTEIN SPECIES PREDOMINATE IN EARLY EMBRYOS AND ARE				
CC	UPREGULATED IN COMMITTED MYOBLASTS AND CHONDROCYTES, BUT				
CC	DOWNREGULATED UPON TERMINAL DIFFERENTIATION. LARGE SPECIES ARE				
CC	DETECTED MAINLY IN ADULT TISSUES AND TERMINALLY DIFFERENTIATED				
CC	CELLS.				
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.				
CC	-!- TISSUE SPECIFICITY: A BROAD PATTERN OF EXPRESSION OBSERVED IN				
CC	TISSUES OF DIVERSE ORIGINS, SUCH AS CARTILAGE, LIVER, BRAIN, LUNG,				
CC	HEART AND SKELETAL MUSCLE. THERE ARE 2 DISTINCT PROTEIN SPECIES:				
CC	THE LARGER ONE (230-250KD) IS FOUND MAINLY IN ADULT BRAIN, LUNG				
CC	AND HEART, AND THE SMALLER ONE (180-190KD) PREDOMINATES IN EARLY				
CC	EMBRYONIC TISSUES.				
CC	-!- DOMAIN: ASN AT POSITION 47 OF THE HOMEBOX MAY PARTICIPATE IN				
CC	REGULATING DNA-BINDING ACTIVITY BY PROMOTING HOMO- AND				
CC	HETERODIMERIZATION.				
CC	-!- ALTERNATIVE PRODUCTS: THOUGH IT IS UNKNOWN WHETHER THE				
CC	HETEROGENEITY OF CLOX PROTEINS IS DUE TO DIVERSITY AT THE				
CC	TRANSCRIPTIONAL LEVEL AND/OR POST-TRANSLATIONAL MODIFICATIONS,				
CC	THERE IS EVIDENCE THAT A NUMBER OF VARIANTS ARE PRODUCED BY				
CC	ALTERNATIVE SPLICING.				
CC	-!- SIMILARITY: CONTAINS THREE CUT-REPEATS AND A HOMEBOX DOMAIN				
CC	SIMILAR TO THOSE OF DROSOPHILA CUT PROTEIN. CDP MAY BE THE				
CC	MAMMALIAN COUNTERPART OF CUT.				
DR	EMBL; X69017; G202; -.				

Search completed: Mon Feb 23 11:02:49 1998
Job time : 27 secs.

WATERMAN
***** (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch_nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Feb 23 21:02:30 1998; MasPar time 235.86 Seconds
981.424 Million cell updates/sec

* Tabular output not generated.

Title: >US-08-836-075A-1
Description: (1-327) from US08836075A.seq
Perfect Score: 317
N.A. Sequence: 1 ATGAGCAGCAATCTTAACCC.....CTGGCGCAATTTGGGCGCC 327
Comp: TACTCGTCTTAGGATTGG.....GAGCGCGTTAACC CGCGGG

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0; Query 0
Searched: 1033316 seqs, 317879456 bases x 2

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: n-pending
1: PCT91 2: PCT92 3: PCT93 4: PCT94 5: PCT95 6: PCT96 7: PCT97
8: U75 9: U76 10: U77A 11: U77B 12: U78A 13: U78B 14: U79A
15: U79B 16: U800 17: U801 18: U802 19: U803 20: U804 21: U805
22: U806 23: U807 24: U808 25: U809 26: U810 27: U811 28: U812
29: U813 30: U814 31: U815 32: U816 33: U817 34: U818 35: U819
36: U820 37: U821 38: U822 39: U823 40: U824 41: U825 42: U826
43: U827 44: U828 45: U829 46: U830 47: U831 48: U832 49: U833
50: U834 51: U835 52: U836 53: U837 54: U838 55: U839 56: U840
57: U841 58: U842 59: U843 60: U844 61: U845 62: U846 63: U847
64: U848 65: U849 66: U85 67: U86 68: U87A 69: U87B 70: U87C
71: U88A 72: U88B 73: U88C 74: U89 75: U6000 76: U6001
77: U6002A 78: U6002B 79: U6003A 80: U6003B 81: U6004A
82: U6004B 83: U6005 84: U6006 85: NEWP 86: NEWU6 87: NEWU8

Statistics: Mean 8.813; Variance 3.802; scale 2.318

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Match	Length	ID	Description	Pred. No.
No.	Score					

1	317	100.0	327	87	US-08-836-	Sequence 1, Applicatio	1.88e-275
2	305	96.2	1645	9	US-07-697-	Sequence 68, Applicati	1.07e-263
3	285	89.9	573	45	US-08-290-	Sequence 114, Applicat	3.89e-244
4	285	89.9	573	45	US-08-290-	Sequence 113, Applicat	3.89e-244
5	283	89.3	327	87	US-08-836-	Sequence 5, Applicatio	3.49e-242
6	283	89.3	573	45	US-08-290-	Sequence 110, Applicat	3.49e-242
7	282	89.0	549	60	US-08-441-	Sequence 62, Applicati	3.31e-241
8	282	89.0	549	38	US-08-221-	Sequence 62, Applicati	3.31e-241
9	282	89.0	549	60	US-08-441-	Sequence 62, Applicati	3.31e-241
10	281	88.6	384	87	US-08-942-	Sequence 18, Applicati	3.14e-240
11	281	88.6	384	66	US-08-537-	Sequence 18, Applicati	3.14e-240
12	281	88.6	573	45	US-08-290-	Sequence 117, Applicat	3.14e-240
13	231	88.6	573	45	US-08-290-	Sequence 111, Applicat	3.14e-240
14	281	88.6	573	45	US-08-290-	Sequence 118, Applicat	3.14e-240
15	281	88.6	573	45	US-08-290-	Sequence 115, Applicat	3.14e-240
16	281	88.6	686	61	US-08-452-	Sequence 25, Applicati	3.14e-240
17	281	88.6	1539	63	US-08-470-	Sequence 17, Applicati	3.14e-240
18	281	88.6	1539	10	US-07-712-	Sequence 17, Applicati	3.14e-240
19	281	88.6	1863	10	US-07-712-	Sequence 14, Applicati	3.14e-240
20	281	88.6	1863	63	US-08-470-	Sequence 14, Applicati	3.14e-240
21	281	88.6	1880	11	US-07-790-	Sequence 28, Applicati	3.14e-240
22	281	88.6	2540	13	US-07-866-	Sequence 2, Applicatio	3.14e-240
23	281	88.6	2540	41	US-08-255-	Sequence 2, Applicatio	3.14e-240
24	281	88.6	9472	31	US-08-150-	Sequence 96, Applicati	3.14e-240
25	280	88.3	2433	67	US-08-612-	Sequence 49, Applicati	2.97e-239
26	279	88.0	345	9	US-07-635-	Sequence 5, Applicatio	2.82e-238
27	279	88.0	345	54	US-08-384-	Sequence 7, Applicatio	2.82e-238
28	279	88.0	573	45	US-08-290-	Sequence 112, Applicat	2.82e-238
29	279	88.0	573	45	US-08-290-	Sequence 124, Applicat	2.82e-238
30	279	88.0	573	45	US-08-290-	Sequence 108, Applicat	2.82e-238
31	279	88.0	1167	9	US-07-635-	Sequence 7, Applicatio	2.82e-238
32	279	88.0	1167	54	US-08-383-	Sequence 7, Applicatio	2.82e-238
33	279	88.0	1167	54	US-08-384-	Sequence 9, Applicatio	2.82e-238
34	279	88.0	1167	48	US-08-324-	Sequence 9, Applicatio	2.82e-238
35	279	88.0	1499	48	US-08-324-	Sequence 3, Applicatio	2.82e-238
36	279	88.0	1665	12	US-07-819-	Sequence 1, Applicatio	2.82e-238
37	279	88.0	6039	48	US-08-324-	Sequence 11, Applicati	2.82e-238
38	279	88.0	6039	11	US-07-769-	Sequence 11, Applicati	2.82e-238
39	279	88.0	9030	9	US-07-635-	Sequence 11, Applicati	2.82e-238
40	279	88.0	9030	25	US-08-099-	Sequence 13, Applicati	2.82e-238
41	279	88.0	9030	11	US-07-769-	Sequence 13, Applicati	2.82e-238
42	279	88.0	9030	54	US-08-384-	Sequence 13, Applicati	2.82e-238
43	279	88.0	9030	11	US-07-769-	Sequence 13, Applicati	2.82e-238
44	279	88.0	9416	9	US-07-635-	Sequence 1, Applicatio	2.82e-238
45	279	88.0	9416	54	US-08-384-	Sequence 1, Applicatio	2.82e-238

ALIGNMENTS

RESULT 1
ID US-08-836-075A-1 STANDARD; DNA; UNC; 327 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 1, Application US/08836075A.
CC Sequence 1, Application US/08836075A
CC GENERAL INFORMATION:
CC APPLICANT: MAERTENS, GEERT
CC APPLICANT: STUYVER, LIEVEN
CC TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
CC TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIA
GNOSTIC
CC TITLE OF INVENTION: AGENTS
CC NUMBER OF SEQUENCES: 207

CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836, 075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29,775
CC REFERENCE/DOCKET NUMBER: INNS:1004
CC INFORMATION FOR SEQ ID NO: 1:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 327 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
SQ Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;

Query Match 100.0%; Score 317; DB 87; Length 327;
Best Local Similarity 100.0%; Pred. No. 1.88e-275;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
QY 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 61 GGGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTCCGGCGGAGG 120
|||||
QY 61 GGGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTCCGGCGGAGG 120
|||||
Db 121 GGGCCAGGNNGGGTGTGGCGCGGACTAGGAAGACTTCCGAGCGGTGCTGGGTGCTGGC 180
|||||
QY 121 GGGCCAGGNNGGGTGTGGCGCGGACTAGGAAGACTTCCGAGCGGTGCTGGGTGCTGGC 180
|||||
Db 181 AGGCGACAGCCTATCCCAAGGCTGCGGCGGAGGAGGCTTCCGGCTGAGCGCGG 240
|||||
QY 181 AGGCGACAGCCTATCCCAAGGCTGCGGCGGAGGAGGCTTCCGGCTGAGCGCGG 240
|||||
Db 241 TATCCTTGGCCCTCTATGGCAATGAGGGCTGGCGGTGGCGGCTGGCTCTGTCCTCC 300
|||||
QY 241 TATCCTTGGCCCTCTATGGCAATGAGGGCTGGCGGTGGCGGCTGGCTCTGTCCTCC 300
|||||
Db 301 CGCGGCTCTCGGCCCAATTTGGGGCCCC 327

QY 301 CGCGGCTCTCGGCCCAATTTGGGGCCCC 327
|||||
RESULT 2
ID US-07-697-326A-68 STANDARD; DNA; UNC; 1645 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 68, Application US/07697326A.
CC Sequence 68, Application US/07697326A
CC GENERAL INFORMATION:
CC APPLICANT: Cha, Tai-An
CC APPLICANT: Beall, Eileen
CC APPLICANT: Irvine, Bruce
CC APPLICANT: Kolberg, Janice
CC APPLICANT: Urdea, Michael S.
CC TITLE OF INVENTION: POLYNUCLEOTIDE PROBES USEFUL FOR
CC TITLE OF INVENTION: SCREENING FOR HEPATITIS C VIRUS
CC NUMBER OF SEQUENCES: 68
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Morrison & Foerster
CC STREET: 545 Middlefield Road, Suite 200
CC CITY: Menlo Park
CC STATE: California
CC COUNTRY: USA
CC ZIP: 94025
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.25
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/07/697, 326A
CC FILING DATE: 19910508
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Reed, Dianne E.
CC REGISTRATION NUMBER: 31,292
CC REFERENCE/DOCKET NUMBER: 2300-0207.00
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-327-7250
CC TELEFAX: 415-327-2951
CC TELEX: 706141
CC INFORMATION FOR SEQ ID NO: 68:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 1645 base pairs
CC TYPE: NUCLEIC ACID
CC STRANDEDNESS: single
CC TOPOLOGY: linear
SQ Sequence 1645 BP; 229 A; 373 C; 384 G; 249 T; 410 other;

Query Match 96.2%; Score 305; DB 9; Length 1645;
Best Local Similarity 80.1%; Pred. No. 1.07e-263;
Matches 262; Conservative 16; Mismatches 16; Indels 0; Gaps 0;
Db 342 A-GAGCAGCCTATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCACAG 401
|||||
QY 1 ATGACGACGAATCTTAACCTCAAGAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 402 GAGCTYAAGTCTCCGGGCGGCGGCTGAGTCTGTGGTGGAGTTTACCTGTGGCGGAGG 461
|||||
QY 61 GGGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTGGCGGAGG 120
|||||
Db 462 GGGCCYAGRTTGGGTGTGGCGGCGGCGGCTGAGTCTTCCGAGCGGTGCGCAACCTCGCGW 521

QY	121	GGCCCCGAGNNGGTGTGCGCGGCACCTAGGAACACTTCCGACGGCTCACACCTGCTGGC	180
Db	522	AGRCGWCARCCTATCCCYAAGCGDCGYCRRCCGACGGAGYAGRCCTGGCGCYAGCCYGGG	581
QY	181	AGGCGACAGCCTATATCCCAAGGCTGTCGCGYCCGAGGACAGGTCTTGGGCTCAGCCCGGG	240
Db	582	TATCCYTGCCCCCTCTATGGYAAAGAGGYWKSGRGTGGCGRGAATGGCTCTCTGTCTMCCC	641
QY	241	TATCCTTGCCCCCTCTATGGCAATAGGCTCGCGGGTGGCGGNTGCTCTCTGTCTCCCC	300
Db	642	CGYGGYTCYCGGCGCTATYTGCGGCCCCY	668
QY	301	CGCGGCTCTCGGCGCCAAATTGGGCGCCC	327

RESULT 3

ID US-08-290-665A-114 STANDARD: DNA: UNC: 573 BP.

AC xxxxxx
DT 01-JAN-1900
DE Sequence 114, Application US/08290665A.
CC Sequence 114, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PORCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE I AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263

RESULT

ID	US-08-290-665A-113 STANDARD; DNA; UNC; 573 BP.
AC	xxxxxx
DT	01-JAN-1900
DE	Sequence 113, Application US/08290665A.
CC	Sequence 113, Application US/08290665A
CC	GENERAL INFORMATION:
CC	APPLICANT: BUKH, J., MILLER, R. H. AND
CC	APPLICANT: PURCELL, R.H.
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC	TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC	NUMBER OF SEQUENCES: 263
CC	CORRESPONDENCE ADDRESS:
CC	ADDRESSEE: MORGAN & FINNEGAN
CC	STREET: 345 PARK AVENUE
CC	CITY: NEW YORK
CC	STATE: NEW YORK
CC	COUNTRY: USA
CC	ZIP: 10154
CC	COMPUTER READABLE FORM:
CC	MEDIUM TYPE: FLOPPY DISK
CC	COMPUTER: IBM PC COMPATIBLE
CC	OPERATING SYSTEM: PC-DOS/MS-DOS
CC	SOFTWARE: WORDPERFECT 5.1
CC	CURRENT APPLICATION DATA:
CC	APPLICATION NUMBER: US/08/290,665A
CC	FILING DATE: 15-AUG-1994
CC	CLASSIFICATION: 435
CC	ATTORNEY/AGENT INFORMATION:
CC	NAME: RICHARD W. BORK

```
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 113:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: P10
SQ Sequence 573 BP; 96 A; 185 C; 179 G; 113 T; 0 other;

Query Match      89.9%; Score 285; DB 45; Length 573;
Best Local Similarity 90.8%; Pred. No. 3,89e-244;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAAACCTCAAGAAACCAACAACTAACACCAACCGCGCCACAG 60
    |||||||
Qy 1 ATGAGCAGCAATCTTAAACCTCAAGAAACCAACAACTAACACCAACCGCGCCCTCAK 60
    |||||||

Db 61 GAGCTCAAGTTCCGGCGGTGTCAGATCTTGTGGAGTTTACCTGTCGGCGCGAGG 120
    |::| |||||
Qy 61 GGGTNNNNNNCCGGGTGGCGGTGAGATCTTGTGGAGTTTACCTGTCGGCGCGAGG 120
    |||||||

Db 121 GCGCCAGAGTTGGTGTGCGCGGACTAGGAAGACTTCCAGCGGTGCGCAACCTCGTGA 180
    |||||||
Qy 121 GCGCCAGAGTTATCCCAAGGCTGCGGCGGAGGAGTCTTGGGCTCAGCGCGGG 240
    |||||||

Db 181 AGCGCACAACTATCCCAAGGCTCGCGCGGAGGAGGCTGGGCTCAGCGCGGG 240
    |||||||
Qy 181 AGCGCACAGCTATCCCAAGGCTGCGGCGGAGGAGTCTTGGGCTCAGCGCGGG 240
    |||||||

Db 241 TACCTTGGCGGCTCTATGGAATGAGGCTTGGGTGGCGAGGATGGCTCTGTCAACC 300
    |||||||
Qy 241 TATCTTGGCGGCTCTATGGAATGAGGCTTGGGTGGCGAGGATGGCTCTGTCAACC 300
    |||||||

Db 301 CGTGGCTCTGGGCTAGTTGGGCGGCC 327
    |||||||
Qy 301 CGCGGCTCTGGGCGCAATTTGGGCGGCC 327
    |||||||

RESULT 5
ID US-08-836-075A-5 STANDARD; DNA; UNC; 327 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 5, Application US/08836075A.
CC Sequence 5, Application US/08836075A
CC GENERAL INFORMATION:
CC APPLICANT: MAERTENS, GEERT
CC APPLICANT: STUYVER, LIEVEN
CC TITLE OF INVENTION: NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES
CC TITLE OF INVENTION: AND THEIR USE AS PROPHYLACTIC, THERAPEUTIC AND DIA
    GNOTIC
CC TITLE OF INVENTION: AGENTS
CC NUMBER OF SEQUENCES: 207
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: ARNOLD, WHITE & DURKEE
CC STREET: P.O. BOX 4433
CC CITY: HOUSTON
```

```
CC STATE: TEXAS
CC COUNTRY: USA
CC ZIP: 77210-4433
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Microsoft Word 6.0 / ASCII text output
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/836, 075A
CC FILING DATE: 21 Apr 1997
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: PCT/EP95/04155
CC FILING DATE: 23 Oct 1995
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 94870166.9
CC FILING DATE: 21 Oct 1994
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: EP 95870076.7
CC FILING DATE: 28 Jun 1995
CC ATTORNEY/AGENT INFORMATION:
CC NAME: KAMMERER, PATRICIA A.
CC REGISTRATION NUMBER: 29, 775
CC REFERENCE/DOCKET NUMBER: INNS:004
CC INFORMATION FOR SEQ ID NO: 5:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 327 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: cDNA
CC HYPOTHETICAL: NO
CC ANTI-SENSE: NO
SQ Sequence 327 BP; 62 A; 105 C; 103 G; 54 T; 3 other;

Query Match      89.3%; Score 283; DB 87; Length 327;
Best Local Similarity 90.3%; Pred. No. 3.49e-242;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAAACCTCAAGAAACCAACAACTAACACCAACCGCGCCACAG 60
    |||||||
Qy 1 ATGAGCAGCAATCTTAAACCTCAAGAAACCAACAACTAACACCAACCGCGCCCTCAK 60
    |||||||

Db 61 GAGCTCAAGNTCCCGGGTGGTGCAGATCGTTGTGGAGTTTACCTGTTGGCGCGAGG 120
    |::| |
Qy 61 GGGTNNNNNNCCGGGTGGCGGTGAGATCGTTGTGGAGTTTACCTGTTGGCGCGAGG 120
    |||||||

Db 121 GCGCCAGAGTTGGTGTGCGCGGCAACCAAGAAAGCTTCCAGCGGTGCGAGCTCTGTAC 180
    |||||||
Qy 121 GCGCCAGAGNNGGTTGTCGCGCGGACTAGGAAGACTTCCAGCGGTGAGAACTCTGTGGC 180
    |||||||

Db 181 AGCGCACAGCTATTTCCTAAGGCTCGCGAGTCCGATGGCAGNNCTGGGCTCAGCGAGG 240
    |||||||
Qy 181 AGCGCACAGCTATTTCCTAAGGCTCGCGAGGCTCGYCGYCCGAGGGGAGGCTCTGGGCTCAGCGCGG 240
    |||||||

Db 241 CATCCCTGGCCCTCTATGGCAATGAGGCTCGCGGATGGCGGGGATGGCTCTGTGCCCC 300
    |||||
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTCGCGGATGGCGGGGATGGCTCTGTGCCCC 300
    |||||

Db 301 CGCGGCTCTCGGCTCAGTTGGGCGGCC 327
    |||||||
Qy 301 CGCGGCTCTCGGCGCAATTTGGGCGGCC 327
    |||||||
```

```

RESULT 6
ID US-08-290-665A-110 STANDARD; DNA; UNC; 573 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 110, Application US/08290665A.
CC Sequence 110, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 110:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: hom sapiens
CC INDIVIDUAL ISOLATE: S45
SQ Sequence 573 BP; 95 A; 189 C; 177 G; 112 T; 0 other;

Query Match 89.3%; Score 283; DB 45; Length 573;
Best Local Similarity 90.5%; Pred. No. 3.49e-242;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCTTAACCTCAAGACAAACCAACCTTAACACCAACCGCGCCACAG 60
|: |||
Qy 1 ATGACGACGAATCTTAACCTCAAGACAAACCAACCTTAACACCAACCGCGCCACAG 60

Db 61 GAGGTCAAGTTCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTTCGCCGCCAGG 120
|: |||
Qy 61 GGSGTNNNNNNCCGGGTGGCGGTGAGTCTGTGGTGGAGTTTACCTGTTCGCCGCCAGG 120

Db 121 GGCCCCAGGTGGGTGGCGCGCGCTAGGACACTCCGACGGGTCAACACTCGTGG 180
|: |||
```

```

Qy 121 GGCCCCAGGNNGGTTGTGCGCGCGCACTAGCAAGACTTCCGACGGGTCAACACTCGTGGC 180
Db 181 CGGCGACAACCTATCCCCAAGGCTCGCGCGCGCGGAGGCGAGGCGCTGGGCGCAGCCGGG 240
|: |||
Qy 181 AGGCGACAGCCTATCCCCAAGGCTCGCGGCGGAGGCGAGGCTCTGGGCTCAGCCCGGG 240
Db 241 CATCTTGGCCCTCTATGGCAATGAGGCTTGGGTTGGGCGGAGGATGGCTCTGTGCACCC 300
|: |||
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTTGGGTTGGGCGGAGGATGGCTCTGTGCACCC 300
Db 301 CGTGCTCGCGCTCTAGTTGGGCGCCC 327
|: |||
Qy 301 CGGCGCTCGCGCTCTAGTTGGGCGCCC 327

RESULT 7
ID US-08-441-971-62 STANDARD; DNA; UNC; 549 BP.
AC xxxxxx
DT 01-JAN-1900
DE Sequence 62, Application US/08441971.
CC Sequence 62, Application US/08441971
CC GENERAL INFORMATION:
CC APPLICANT: Tai-An Cha
CC TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR
CC TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
CC NUMBER OF SEQUENCES: 147
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Wolf, Greenfield & Sacks, P.C.
CC STREET: 600 Atlantic Avenue
CC CITY: Boston
CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 5.25 inch
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: MS-DOS Version 3.3
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,971
CC FILING DATE: 16-MAY-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US/08/221,653
CC FILING DATE:
CC APPLICATION NUMBER: US/07/881,528
CC FILING DATE:
CC APPLICATION NUMBER: 07/697,326
CC FILING DATE: 8 May 1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Janiuk, Anthony J.
CC REGISTRATION NUMBER: 29,809
CC REFERENCE/DOCKET NUMBER: C0772/7000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 720-3500
CC TELEFAX: (617) 720-2441
CC TELEX: EZEKIEL
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 549 nucleotides
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA
```


CC	ORIGINAL SOURCE:	
CC	INDIVIDUAL ISOLATE: spl	
SQ	Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;	
	Query Match	89.0%; Score 282; DB 60; Length 549;
	Best Local Similarity	90.5%; Pred. No. 3.31e-241;
	Matches	295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db	1	ATGAGCAGCAATCCTAAACCTCAAGAAAGAAACCAAACTAAACCAACGCGCGCCACAG 60
Qy	1	ATGAGCAGCAATCCTAAACCTCAAGAAAGAAACCAAACTAAACCAACGCGCGCCCTAK 60
Db	61	GAGCTCAAGTCCCGGGCGGTGTCAGACTGTTGGTGGAGTTTACCTGTCGCGCGCAGG 120
Qy	61	GGSGTNNNNNNCCGGGTGGCGGTTCAGACTGTTGGTGGAGTTTACCTGTCGCGCGCAGG 120
Db	121	GGCCCAAGTTGGGTGTGGCGCGCACTAGCAAGACTCCCGAGCGGTCCGCAACCTCGTGA 180
Qy	121	GGCCCAAGNNGGTGTGGCGCGCACTAGCAAGACTCCCGAGCGGTTCGCAACCTCGTGGC 180
Db	181	AGGCGCAACACTATCCCAAGGCTCGCGCGCCGCGAGGCGCTGGGCTACGCGCGGG 240
Qy	181	AGGCGAGAGCTATCCCAAGGCTGCGGCGGAGGCGAGGCTCTGGGCTCAGCGCGGG 240
Db	241	TATCCTTGGCCCTCTATGCCAATCAGGGTCTCGGGTGGCAGGATGGCTCTGTCAACC 300
Qy	241	TATCCTTGGCCCTCTATGGCAATGAGGGCTCGGGTGGCGGGATGGCTCTGTCTCCCC 300
Db	301	CGCGGCTCTCGGCTAGCTGGGCGCC 326
Qy	301	CGCGGCTCTCGGCGCAATTGGGCGCC 326
RESULT	8	
ID	US-08-221-653-62	STANDARD; DNA; UNC; 549 BP.
AC	xxxxxx	
DE	01-JAN-1900	
DT	Sequence 62,	Application US/08221653.
CC	Sequence 62,	Application US/08221653
CC	GENERAL INFORMATION:	
CC	APPLICANT:	Tai-An Cha
CC	TITLE OF INVENTION:	HCV GENOMIC SEQUENCES FOR
CC	TITLE OF INVENTION:	DIAGNOSTICS AND THERAPEUTICS
CC	NUMBER OF SEQUENCES:	147
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	Wolf, Greenfield & Sacks, P.C.
CC	STREET:	600 Atlantic Avenue
CC	CITY:	Boston
CC	STATE:	Massachusetts
CC	COUNTRY:	USA
CC	ZIP:	02210
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	Diskette, 5.25 inch
CC	COMPUTER:	IBM compatible
CC	OPERATING SYSTEM:	MS-DOS Version 3.3
CC	SOFTWARE:	WordPerfect 5.1
CC	CURRENT APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/08/221,653
CC	FILING DATE:	
CC	CLASSIFICATION:	435
CC	PRIOR APPLICATION DATA:	
CC	APPLICATION NUMBER:	US/07/881,528
CC	FILING DATE:	
CC	APPLICATION NUMBER:	07/697,326

CC	FILING DATE: 8 May 1991	
CC	ATTORNEY/AGENT INFORMATION:	
CC	NAME: Janiuk, Anthony J.	
CC	REGISTRATION NUMBER: 29,809	
CC	REFERENCE/DOCKET NUMBER: C0772/7000	
CC	TELECOMMUNICATION INFORMATION:	
CC	TELEPHONE: (617) 720-3500	
CC	TELEFAX: (617) 720-2441	
CC	TELEX: EZEKIEL	
CC	INFORMATION FOR SEQ ID NO: 62:	
CC	SEQUENCE CHARACTERISTICS:	
CC	LENGTH: 549 nucleotides	
CC	TYPE: nucleic acid	
CC	STRANDEDNESS: single	
CC	TOPOLOGY: linear	
CC	MOLECULE TYPE: DNA	
CC	ORIGINAL SOURCE:	
CC	INDIVIDUAL ISOLATE: spl	
CC	Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;	
CC	Query Match 89.0%; Score 282; DB 38; Length 549;	
CC	Best Local Similarity 90.5%; Pred.No. 3.31e-241;	
CC	Matches 295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;	
DB	1 ATGAGCAGCATCTTAACCTCAAGAAAAACCAACCTTAACCAACCAACGCGCGCCACAG 60	
QY	1 ATGAGCAGCATCTTAACCTCAAGAAAAACCAACCTTAACCAACCAACGCGCGCCACAG 60	
DB	61 GAGCTCAAGTTCCCGCGCGGTGTCAGATGTTGTGCGAGTTTACCTGTTGCCGCGCAGG 120	
QY	61 GSGGTNNNNNNCCGGGTGGCGGTGTCAGATGTTGTGGAGTTTACCTGTTGCCGCGCAGG 120	
DB	121 GGCCCGAGGTTGGGTGCGCGCGACTAGGAAGACTTCCGACGCGTCCGAACCTGCTGGA 180	
QY	121 GGCCCGAGNNGGGTGCGCGCGACTAGGAAGACTTCCGACGCGTCCGAACCTGCTGGC 180	
DB	181 AGCGGACAACTATCCCAAGGCTCGCGCGCCCGCAGGCGAGGCTGGGCTAGCCCCGG 240	
QY	181 AGCGGACAGCTATCCCAAGGCTCGYCGGCGCAGGCGAGGCTCTGGGCTCAGCCCCGG 240	
DB	241 TATCCTTGGCCCTCTATGGCAATGAGGGTCTGGGGTGGGCGAGAGTGGCTCTGTACCC 300	
QY	241 TATCCTTGGCCCTCTATGGCAATGAGGGCTGGGGTGGGCGGNTGGCTCTGTCCCCC 300	
DB	301 CGCGGCTCTCGGCTAGCTGGCGGCC 326	
QY	301 CGCGGCTCTCGGCGCAATTGGGGCC 326	
RESULT	9	
ID	US-08-141-970-62 STANDARD; DNA; UNC; 549 BP.	
AC	xxxxxx	
DT	01-JAN-1900	
DE	Sequence 62, Application US/08441970.	
CC	Sequence 62, Application US/08441970	
CC	GENERAL INFORMATION:	
CC	APPLICANT: Tai-An Cha	
CC	TITLE OF INVENTION: HCV GENOMIC SEQUENCES FOR	
CC	TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS	
CC	NUMBER OF SEQUENCES: 147	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: Wolf, Greenfield & Sacks, P.C.	
CC	STREET: 600 Atlantic Avenue	
CC	CITY: Boston	

CC STATE: Massachusetts
CC COUNTRY: USA
CC ZIP: 02210
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette, 5.25 inch
CC COMPUTER: IBM compatible
CC OPERATING SYSTEM: MS-DOS Version 3.3
CC SOFTWARE: WordPerfect 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/441,970
CC FILING DATE: 16-MAY-1995
CC CLASSIFICATION: 536
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 07/881,528
CC FILING DATE: 08-MAY-1992
CC APPLICATION NUMBER: 07/697,326
CC FILING DATE: 8 May 1991
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Janiak, Anthony J.
CC REGISTRATION NUMBER: 29,809
CC REFERENCE/DOCKET NUMBER: C0772/7000
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (617) 720-3500
CC TELEFAX: (617) 720-2441
CC TELEX: EZEKIEL
CC INFORMATION FOR SEQ ID NO: 62:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 549 nucleotides
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: DNA
CC ORIGINAL SOURCE:
CC INDIVIDUAL ISOLATE: spl
SQ Sequence 549 BP; 92 A; 175 C; 173 G; 109 T; 0 other;

Query Match 89.0%; Score 282; DB 60; Length 549;
Best Local Similarity 90.5%; Pred. No. 3.31e-241;
Matches 295; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60
61 GAGCTCAAGTTCCGGCGGCTGTCAGATGTTGGTGGAGTTTACCTGTCGCCGCCAGG 120
|: || |||| || |||||
Qy 61 GGGTNNNNNNCGGGTGGCGGTGAGTCTGTTGGAGTTTACCTGTCGCCGCCAGG 120
Db 121 GGGCCAGCTTGGCTGTCGGCGGCTAGGACACTCCGAGCGGTCCGACACTCTGGA 180
|||||
Qy 121 GGGCCAGCTTGGCTGTCGGCGGCTAGGACACTCCGAGCGGTCCGACACTCTGTCG 180
Db 181 AGCGCAGCACTATCCCAAGGCTCGCGCGGCGGCGGCGGCTGGGCTCAGCGCGG 240
|||||
Qy 181 AGCGCAGCACTATCCCAAGGCTCGCGCGGCGGCGGCGGCTGGGCTCAGCGCGG 240
Db 241 TATCCTTGGCCCTCTATGGCAATGAGGGTCTGGGGTGGGCGAGGATGGCTCTGTCA 300
|||||
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGGTCTGGGGTGGGCGAGGATGGCTCTGT 300
Db 301 CGGGCTCTGGGCTAGTGGGGGCC 326
Qy 301 CGGGCTCTGGGCGCAATTGGGGGCC 326

RESULT 10
ID US-08-942-483-18 STANDARD; DNA; UNC; 384 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 18, Application US/08942483.
CC Sequence 18, Application US/08942483
CC GENERAL INFORMATION:
CC APPLICANT: Joong, Myung
CC APPLICANT: Choi, Deog Young
CC APPLICANT: Kim, Chun Hyung
CC APPLICANT: Yang, Jae Young
CC APPLICANT: Kim, In Soo
CC APPLICANT: Choi, Dong Seob
CC TITLE OF INVENTION: DIAGNOSTIC KIT AND METHOD
CC TITLE OF INVENTION: FOR THE SIMULTANEOUS DIAGNOSIS OF HEPATITIS B AND
CC NUMBER OF SEQUENCES: 35
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: PENNIE & EDMONDS LLP
CC STREET: 1155 Avenue of the Americas
CC CITY: New York
CC STATE: NY
CC COUNTRY: U.S.A
CC ZIP: 10036-2711
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSeq Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/942,483
CC FILING DATE: 02-OCT-1997
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/537,814
CC FILING DATE: 24-OCT-1995
CC APPLICATION NUMBER: KR-93-7231
CC FILING DATE: 28-APR-1993
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Jones III, Harry C.
CC REGISTRATION NUMBER: 20,280
CC REFERENCE/DOCKET NUMBER: 8512-036-999
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 790-9090
CC TELEFAX: (212) 869-9741
CC TELEX:
CC INFORMATION FOR SEQ ID NO: 18:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 384 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC FEATURE:
CC OTHER INFORMATION: KHCV CORE14, Fig. 1
SQ Sequence 384 BP; 76 A; 120 C; 121 G; 67 T; 0 other;

Query Match 88.6%; Score 281; DB 87; Length 384;
Best Local Similarity 90.5%; Pred. No. 3.14e-240;
Matches 296; Conservative 3; Mismatches 28; Indels 0; Gaps 0;
Db 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTAK 60


```
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 117:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND3
CC Sequence 573 BP; 100 A; 184 C; 174 G; 115 T; 0 other;

Query Match      88.6%; Score 281; DB 45; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db 61 GAGCTCAAGTTCGCGGGGGTGCAGACTGTTGGTGGAGTTTACCTGTGCGCGCGCAGG 120
Qy 61 GSGTNNNNNNCCGGGTGGCGGTGCAGACTGTTGGTGGAGTTTACCTGTGCGCGCGCAGG 120

Db 121 GCGCCAGGTTGGGTGGCGCGGACTAGGAGACTTCCGAGCGGTGCCAACCTCGTGA 180
Qy 121 GCGCCAGGNNGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCACACCTCGTGGC 180

Db 181 AGCGCACACCTATCCCAAGGCTCGCGCGGAGGAGGCTAGGGCTTGGGCTCAGCCCGG 240
Qy 181 AGCGCACAGCCTATCCCAAGGCTCGCGCGGAGGAGGCTAGGGCTTGGGCTCAGCCCGG 240

Db 241 TACCTTGGCCCTCTATGCAATCAGGGCTTGGGTGGCGAGATGCTCTGTGACCC 300
Qy 241 TATCTTGGCCCTCTATGCAATCAGGGCTTGGGTGGCGGAGTGGCTCTGTGACCC 300

Db 301 CGCGGTTCTCGGCCTAGTTGGGCCCC 327
Qy 301 CGCGGCTCTCGGCCAATTTGGGCCCC 327

RESULT 13
ID US-08-290-665A-111 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 111, Application US/08290665A.
CC Sequence 111, Application US/08290665A
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R. H. AND
```

```
CC APPLICANT: PORCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/290, 665A
CC FILING DATE: 15-AUG-1994
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 111:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: D1
CC Sequence 573 BP; 96 A; 184 C; 177 G; 116 T; 0 other;

Query Match      88.6%; Score 281; DB 45; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAAAACCAACGTAACACCAACCGCGCCACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db 61 GAGCTCAAGTTCGCGGGGGTGCAGACTGTTGGTGGAGTTTACCTGTGCGCGCGCAGG 120
Qy 61 GSGTNNNNNNCCGGGTGGCGGTGCAGACTGTTGGTGGAGTTTACCTGTGCGCGCGCAGG 120

Db 121 GCGCCAGGTTGGGTGGCGCGGACTAGGAGACTTCCGAGCGGTGCCAACCTCGTGA 180
Qy 121 GCGCCAGGNNGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTGCACACCTCGTGGC 180

Db 181 AGCGCACACCTATCCCAAGGCTCGCGCGGAGGAGGCTAGGGCTTGGGCTCAGCCCGG 240
Qy 181 AGCGCACAGCCTATCCCAAGGCTCGCGCGGAGGAGGCTTGGGCTCAGCCCGG 240

Db 241 TACCTTGGCCCTCTATGCAATCAGGGCTTGGGTGGCGAGATGCTCTGTGACCC 300
Qy 241 TATCTTGGCCCTCTATGCAATCAGGGCTTGGGTGGCGGAGTGGCTCTGTGACCC 300
```


CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:

CC ORGANISM: *homo sapiens*
CC INDIVIDUAL ISOLATE: T10

SQ Sequence 573 BP; 99 A; 181 C; 176 G; 117 T; 0 other;

Query Match 88.6%; Score 281; DB 15; Length 573;
Best Local Similarity 90.2%; Pred. No. 3.14e-240;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGACGACGAACTCTAAACCTCAAGAAACCAACACGTTACACCAACCGCGCCACAG 60
|||

Qy 1 ATGACGACGAACTCTAAACCTCAAGAAACCAACACGTTACACCAACCGCGCCCTCAK 60
|||

Db 61 GACGTCAAGTTCCCGGGGTGTCAGATCGTTGGTGGAGTTTACCTGTCCGCGCAGG 120
|||

Qy 61 GCGTNNNNNNCCGGGTGGCGTCAGATCGTTGGTGGAGTTTACCTGTCCGCGCAGG 120
|||

Db 121 GCGCCAGGTGGGTGTCGCGCGACTAGGAGACTTCCGAGCGGTCCGAACTCGTGA 180
|||

Qy 121 GCGCCAGGNNGGGTGTCGCGCGACTAGGAGACTTCCGAGCGGTCCGAACTCGTGGC 180
|||

Db 181 AGGCGACAGCTATCCCAAGGCTCCGACCGGCGAGGGGCTGGGCTCAGCCCGGG 240
|||

Qy 181 AGGCGACAGCTATCCCAAGGCTCGYCGGTCGAGGGCAGGCTCTGGGCTCAGCCCGGG 240
|||

Db 241 TACCTTGGCCCTCTATGCAATGAGGGCATGGGTGGGAGGATGGCTCTCTCACC 300
|||

Qy 241 TATCCTTGGCCCTCTATGCAATGAGGGCTGGGGTGGGCGGNTGGCTCTCTCACC 300
|||

Db 301 CGTGCTCCCGGCTAGTTGGGCCCC 327
|||

Qy 301 CGCGGCTCTCGGCCAATTGGGCCCC 327
|||

Search completed: Mon Feb 23 21:06:34 1998
Job time : 244 secs.

Best Local Similarity 90.8%; Pred. No. 1.04e-106;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkgfgggqivggvylprtgprlgvratrktsersqrg 60
|||||
Qy 1 MSTNPKPQRTKRNTRNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||

Db 61 rrqpikarrsegrawagpypwplynecgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRKGSRWAQPGYPWPLYNCGCGMAXWLLSPRGSRPNWGP 109
|||||

RESULT 2
ENTRY S41364 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 4, N3)
CONTAINS (fragment)
ORGANISM #formal_name hepatitis C virus
#variety genotype_4, N3
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41364
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

CLASSIFICATION #molecule_type genomic RNA
#residues 1-115 #label VAN
#cross-references EMBL:Z29467
#experimental_source genotype 4, N3

KEYWORDS #superfamily hepatitis C virus genome polyprotein
capsid protein; core protein; polyprotein

FEATURE 1-115
#product core protein #status predicted #label MAT
SUMMARY #length 115 #checksum 8983

Query Match 99.5%; Score 777; DB 9; Length 115;
Best Local Similarity 90.8%; Pred. No. 1.04e-106;
Matches 99; Conservative 1; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkgfgggqivggvylprtgprlgvratrktsersqrg 60
|||||
Qy 1 MSTNPKPQRTKRNTRNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||

Db 61 rrqpikarrsegrawagpypwplynecgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRKGSRWAQPGYPWPLYNCGCGMAXWLLSPRGSRPNWGP 109
|||||

RESULT 3
ENTRY S41363 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 4, N2)
CONTAINS (fragment)
ORGANISM core protein
#variety genotype 4, N2
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41363
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

Accession S41363
#molecule_type genomic RNA
#residues 1-117 #label VAN
#cross-references EMBL:Z29466
#experimental_source genotype 4, N2

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
capsid protein; core protein; polyprotein

KEYWORDS 1-117
#product core protein #status predicted #label MAT
SUMMARY #length 117 #checksum 41

Query Match 99.5%; Score 777; DB 9; Length 117;
Best Local Similarity 91.7%; Pred. No. 1.04e-106;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkgfgggqivggvylprtgprlgvratrktsersqrg 60
|||||
Qy 1 MSTNPKPQRTKRNTRNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||

Db 61 rrqpikarrsegrawagpypwplynecgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRKGSRWAQPGYPWPLYNCGCGMAXWLLSPRGSRPNWGP 109
|||||

RESULT 4
ENTRY S41341 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 1, N1)
CONTAINS (fragment)
ORGANISM core protein
#formal_name hepatitis C virus
#variety genotype 1, N1
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41341
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.

CLASSIFICATION #molecule_type genomic RNA
#residues 1-112 #label VAN
#cross-references EMBL:Z29444
#experimental_source genotype 1, N1

KEYWORDS #superfamily hepatitis C virus genome polyprotein
capsid protein; core protein; polyprotein

FEATURE 1-112
#product core protein #status predicted #label MAT
SUMMARY #length 112 #checksum 9

Query Match 99.2%; Score 775; DB 9; Length 112;
Best Local Similarity 89.9%; Pred. No. 2.23e-106;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnbpqrktkrntnrpmdvkgfgggqivggvylprtgprlgvratrktsersqrg 60
|||||
Qy 1 MSTNPKPQRTKRNTRNRRPXXXXPFGGQIVGGVYLLPRGPRGXGVRATRKTSERSQPRG 60
|||||

Db 61 rrqpikarrsegrawagpypwplynecgwagwllsprgrpswgp 109
|||||
Qy 61 RRQPIPKAXRKGSRWAQPGYPWPLYNCGCGMAXWLLSPRGSRPNWGP 109
|||||

RESULT 5
ENTRY S41345 #type fragment

TITLE genome polyprotein - hepatitis C virus (genotype 1, N3)
 (fragment)
 CONTAINS core protein
 ORGANSIM #formal_name hepatitis C virus
 #variety genotype 1, N3
 DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997
 ACCESSIONS S41345
 REFERENCE S41341
 #authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 #submission submitted to the EMBL Data Library, January 1994
 #description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 #accession S41345
 ##molecule_type genomic RNA
 ##residues 1-115 #label VAN
 ##cross-references EMBL:229448
 ##experimental_source genotype 1, N3
 CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
 KEYWORDS capsid protein; core protein; polyprotein
 FEATURE 1-115
 SUMMARY #product core protein #status predicted #label MAT
 #length 115 #checksum 9357

Query Match 99.2%; Score 775; DB 9; Length 115;
 Best Local Similarity 89.9%; Pred. No. 2.23e-106;
 Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnpkqrktrntnrpqdvkfgggqivggyllprgprlgvratrktsersqprg 60
 |||||
 Qy 1 MSTNPKQRKTRNTRNRPPXXXXPGGGQIVGGVLLPRGPRXGVRATRKTSERSQPRG 60

Db 61 rrcqpkarrpgrtwagpywplgncgcwagwllsprgsrswqp 109
 |||||

Qy 61 RRQPIPKAXRSGRWAQPGYPMPLYGNECGGMAXWLLSPRGRNMGCP 109

RESULT 6

ENTRY

TITLE

CONTAINS S41346 #type fragment
 genome polyprotein - hepatitis C virus (genotype 1, N4)
 (fragment)
 ORGANSIM core protein
 #formal_name hepatitis C virus
 #variety genotype 1, N4
 DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS

REFERENCE

#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
 #submission submitted to the EMBL Data Library, January 1994
 #description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
 #accession S41346

##molecule_type genomic RNA

##residues 1-118 #label VAN

##cross-references EMBL:229449

##experimental_source genotype 1, N4

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
 KEYWORDS capsid protein; core protein; polyprotein

FEATURE

1-118 #product core protein #status predicted #label MAT
 SUMMARY #length 118 #checksum 81

Query Match 99.2%; Score 775; DB 9; Length 118;
 Best Local Similarity 89.9%; Pred. No. 2.23e-106;
 Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnpkqrktrntnrpqdvkfgggqivggyllprgprlgvratrktsersqprg 60
 |||||

Qy 1 MSTNPKQRKTRNTRNRPPXXXXPGGGQIVGGVLLPRGPRXGVRATRKTSERSQPRG 60

Db 61 rrcqpkarrpgrtwagpywplgncgcwagwllsprgsrswqp 109
 |||||

Qy 61 RRQPIPKAXRSGRWAQPGYPMPLYGNECGGMAXWLLSPRGRNMGCP 109

RESULT 7

ENTRY

TITLE

CONTAINS GNVVCH #type complete
 genome polyprotein - hepatitis C virus (strain H)
 capsid protein C; envelope protein M; major envelope protein
 E; nonstructural protein NS1; nonstructural protein NS2;
 nonstructural protein NS3; nonstructural protein NS4a;
 nonstructural protein NS4b; nonstructural protein NS5

ORGANSIM

#note

DATE

31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Feb-1997

ACCESSIONS

REFERENCE

#authors Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.
 #submission submitted to GenBank, July 1992

#description

Genomic structure of the human prototype strain H of
 hepatitis C virus: comparison with American and Japanese
 isolates.

#accession

##molecule_type genomic RNA

##residues 1-3011 #label INC

##cross-references GB:M67463

REFERENCE

#authors

Inchauspe, G.; Zebedee, S.; Lee, D.H.; Sugitani, M.; Nasoff, M.; Prince, A.M.

#journal

#title

Proc. Natl. Acad. Sci. U.S.A. (1991) 88:10292-10296
 Genomic structure of the human prototype strain H of
 hepatitis C virus: comparison with American and Japanese
 isolates.

#cross-references MUID:92052256

#contents

#note

neither amino acid nor nucleotide sequence is given
 #superfamily hepatitis C virus genome polyprotein
 capsid protein; envelope protein; glycoprotein; nonstructural
 protein; polyprotein; transmembrane protein

CLASSIFICATION

KEYWORDS

#product capsid protein C #status predicted #label CPC\
 #product envelope protein M #status predicted #label
 EPV\
 #product major envelope protein E #status predicted
 #label MEE\
 #product nonstructural protein NS1 #status predicted
 #label NS1\
 #product nonstructural protein NS2 #status predicted
 #label NS2\
 #product nonstructural protein NS3 #status predicted
 #label NS3\
 #product nonstructural protein NS4a #status predicted
 #label NS4a\
 #product nonstructural protein NS4b #status predicted
 #label NS4b\
 #product nonstructural protein NS5 #status predicted
 #label NS5\
 #product nonstructural protein NS5 #status predicted
 #label NS5

FEATURE

1-115

116-191

192-389

390-729

730-1006

1007-1615

1616-1862

1863-2013

2014-3011

540,556,576,623	#binding site carbonylrate (Asn)	(covalent)	#status
SUMMARY	#length 640	#checksum 2455	
Query Match	98.3%;	Score 768; DB 9; Length 640;	
Best Local Similarity	89.0%;	Pred. No. 3.25e-105;	
Matches	97; Conservative	2; Mismatches 10; Indels 0; Gaps 0;	
Db	1 mstnkpqrktkntnrpqdvkfpqggqilvgvyllprqprlgvratrktsersqrg	60	
QY	1 MSTNPKPQRKTGNTRRPXXXXXPGGGQIVGGVYLLPRGRPRXGVATRKTSERSQPRG	60	
Db	61 rrcpikparrpgrtwagqypwplvgnegcgwaqllsprgrpswqp	109	
QY	61 RRQIPKAXRGRSWAOPGYPMPLVGNEGCGWAXWLLSPGRSRENWGP	109	
RESULT 11			
ENTRY	S32740	#type fragment	
TITLE	core protein - hepatitis C virus (fragment)		
ORGANISM	#formal name hepatitis C virus		
DATE	19-Mar-1997	#sequence_revision 19-Mar-1997	#text_change
ACCESSIONS	S32740		
REFERENCE	S32740		
#authors	Vasilev, V.B.; Viazov, S.O.; Kotova, E.Y.; Nosikov, V.V.		
#submission	submitted to the EMBL Data Library, April 1993		
#accession	S32740		
#status	preliminary		
#residues	1-189	#label VAS	
#cross-references	EMBL:X71407		
SUMMARY	#length 189	#checksum 7020	
Query Match	98.0%;	Score 765; DB 18; Length 189;	
Best Local Similarity	89.0%;	Pred. No. 1.03e-104;	
Matches	97; Conservative	2; Mismatches 10; Indels 0; Gaps 0;	
Db	1 mstnkpqrktkntnrpqdvkfpqggqilvgvyllprqprlgvratrktsersqrg	60	
QY	1 MSTNPKPQRKTGNTRRPXXXXXPGGGQIVGGVYLLPRGRPRXGVATRKTSERSQPRG	60	
Db	61 rrcpikparrpgrtwagqypwplvgnegcgwaqllsprgrpswqp	109	
QY	61 RRQIPKAXRGRSWAOPGYPMPLVGNEGCGWAXWLLSPGRSRENWGP	109	
RESULT 12			
ENTRY	GNMVC3	#type complete	
TITLE	genome polyprotein - hepatitis C virus (strain HCV-1)		
CONTAINS	capsid, polypeptide C; envelope protein M; major envelope protein E; nonstructural protein NS1; nonstructural protein NS2; nonstructural protein NS3; nonstructural protein NS4a; nonstructural protein NS4b; nonstructural protein NS5		
ORGANISM	#formal name hepatitis C virus		
DATE	30-Sep-1992	#sequence_revision 30-Sep-1992	#text_change
ACCESSIONS	A39166; PQ0403; PQ0404		
REFERENCE	A39166		
#authors	Choo, Q.L.; Richman, K.H.; Han, J.H.; Berger, K.; Lee, C.; Dong, C.; Gallegos, C.; Coit, D.; Medina-Selby, A.; Barr, P.J.; Weiner, A.J.; Bradley, D.W.; Kuo, G.; Houghton, M.		
#journal	Proc. Natl. Acad. Sci. U.S.A. (1991) 88:2451-2455		
#title	Genetic organization and diversity of the hepatitis C virus.		
#cross-references	MUID:91172826		

#accession	A39166
#molecule_type	mRNA
#residues	1-3011 ##label CHO
#cross-references	CB:M62321
REFERENCE	PQ0393
#authors	Chan, S.W.; McOmish, F.; Holmes, E.C.; Dow, B.; Peutherer, J.F.; Follett, E.; Yap, P.L.; Simmonds, P.
#journal	J. Gen. Virol. (1992) 73:1131-1141
#title	Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants.
#accession	PQ0403
#molecule_type	genomic RNA
#residues	1577-1633 ##label CHA
#cross-references	DD83:D10128
#experimental_source	isolates E-b16
#accession	PQ0404
#status	preliminary
#molecule_type	genomic RNA
#residues	1577-1633 ##label CH2
#experimental_source	isolates E-b17
CLASSIFICATION	superfamily hepatitis C virus genome polyprotein
KEYWORDS	capsid protein; envelope protein; glycoprotein; nonstructural protein; polyprotein; transmembrane protein
FEATURE	
1-115	#product capsid protein C #status predicted #label CPC\
116-191	#product envelope protein M #status predicted #label EPM\
192-389	#product major envelope protein E #status predicted #label MEE\
390-729	#product nonstructural protein NS1 #status predicted #label NS1\
730-1006	#product nonstructural protein NS2 #status predicted #label NS2\
1007-1615	#product nonstructural protein NS3 #status predicted #label NS3\
1616-1862	#product nonstructural protein NS4a #status predicted #label N4a\
1863-2013	#product nonstructural protein NS4b #status predicted #label N4b\
2014-3011	#product nonstructural protein NS5 #status predicted #label N5\
196-209, 234, 305, 325, 417, 423, 430, 448, 476, 532, 540, 556, 576, 623, 645, 1213, 1255, 2041, 2071, 2240, 2364, 2550, 2789	
SUMMARY	
Query Match	98.0%; Score 765; DB 4; Length 3011;
Best Local Similarity	88.1%; Pred. No. 1.03e-104;
Matches	96; Conservative 3; Mismatches 10; Indels 0; Gaps 0;
Db	1 mstnptkknkntnrpqdvfgggvggwyllprgprlgvratrktserspqg 60
Qy	1 MSTNPKQRTKGNTRRPPXXXXPGGGVYLLPRGPRGVRATRKTSERSQPRG 60
Db	61 rrrpikarrpegrtwagpywpvlynegcgwagwllprgrspwpg 109
Qv	61 RRPIPKARKEGRWAQGYPPPLYGNEGCGMAXLSPRGRPPWCP 109

RESULT 13

ENTRY S40770 #type complete
TITLE polyprotein precursor - hepatitis C virus
ORGANISM #formal_name hepatitis C virus
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 07-Feb-1997

ACCESSIONS S40770; PC1285
REFERENCE S40770
#authors Okamoto, H.
#submission submitted to the EMBL Data Library, March 1992
#accession S40770
##status preliminary
##molecule_type genomic RNA
##residues 1-3011 ##label OKA
##cross-references EMBL:D10749

REFERENCE PC1284
#authors Okamoto, H.; Okada, S.; Sugiyama, Y.; Yotsumoto, S.; Tanaka, T.; Yoshizawa, H.; Tada, F.; Miyakawa, Y.; Mayumi, M.
#journal J. Exp. Med. (1990) 60:167-177
#title The 5'-terminal sequence of the hepatitis C virus genome.
#accession PC1285
##molecule_type genomic RNA
##residues 1-513 ##label OK2
##experimental_source isolate HC-J1

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS polyprotein
SUMMARY #length 3011 #molecular_weight 327114 #checksum 5911

Query Match 97.7%; Score 763; DB 9; Length 3011;
Best Local Similarity 88.1%; Pred. No. 2.21e-104;
Matches 96; Conservative 2; Mismatches 11; Indels 0; Gaps 0;

Db 1 matnkpqrktkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Qy 1 MSTNPKPQRKTkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Db 61 rrqipkarpgegrtwagpygwpwlygncgcgagwllsprgrpswqp 109

Qy 61 RRQIPKAXRKEGRSWAQGYPWLYGNCGCCGAXWLLSPRGSRNWGP 109

RESULT 14

ENTRY S41344 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 1, N2)
CONTAINS core protein
ORGANISM #formal_name hepatitis C virus
#variety genotype 1, N2
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41344
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession S41344
##molecule_type genomic RNA
##residues 1-115 ##label VAN
##cross-references EMBL:Z29447
##experimental_source genotype 1, N2

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; core protein; polyprotein
FEATURE 1-115
#product core protein #status predicted #label MAT

SUMMARY #length 115 #checksum 9764

Query Match 97.6%; Score 762; DB 9; Length 115;
Best Local Similarity 89.0%; Pred. No. 3.23e-104;
Matches 97; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 1 matnkpqrktkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Qy 1 MSTNPKPQRKTkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Db 61 rrqipkarpgegrtwagpygwpwlygncgcgagwllsprgrpswqp 109

Qy 61 RRQIPKAXRKEGRSWAQGYPWLYGNCGCCGAXWLLSPRGSRNWGP 109

RESULT 15

ENTRY S41343 #type fragment
TITLE genome polyprotein - hepatitis C virus (genotype 1, N11)
CONTAINS core protein
ORGANISM #formal_name hepatitis C virus
#variety genotype 1, N11
DATE 19-May-1994 #sequence_revision 26-Jul-1996 #text_change 16-Feb-1997

ACCESSIONS S41343
REFERENCE S41341
#authors van Doorn, L.J.; Kleter, G.E.M.; Brouwer, J.T.
#submission submitted to the EMBL Data Library, January 1994
#description Analysis of hepatitis C virus genotypes 1 to 5 by LiPA.
#accession S41343
##molecule_type genomic RNA
##residues 1-115 ##label VAN
##cross-references EMBL:Z29446
##experimental_source genotype 1, N11

CLASSIFICATION #superfamily hepatitis C virus genome polyprotein
KEYWORDS capsid protein; core protein; polyprotein
FEATURE 1-115
#product core protein #status predicted #label MAT

SUMMARY #length 115 #checksum 9697

Query Match 96.5%; Score 754; DB 9; Length 115;
Best Local Similarity 89.0%; Pred. No. 6.90e-103;
Matches 97; Conservative 2; Mismatches 10; Indels 0; Gaps 0;

Db 1 matnkpqrktkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Qy 1 MSTNPKPQRKTkrntnrpqdvkfpgggqivggvylprgrprlgvratrktseersqrg 60

Db 61 rrqipkarpgegrtwagpygwpwlygncgcgagwllsprgrpswqp 109

Qy 61 RRQIPKAXRKEGRSWAQGYPWLYGNCGCCGAXWLLSPRGSRNWGP 109

Search completed: Mon Feb 23 11:02:03 1998
Job time : 56 secs.

Database: u-emb151_101
132:VRT 133:PHG 134:PLN1 135:PLN2 136:PRI1 137:PRI2
138:ROO 139:SYN 140:UNA 141:VRL
142:part1 143:part2

Statistics: Mean 9.880; Variance 4.750; scale 2.080

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	317	100.0	327	143	A50346	Sequence 1 from Paten	7,68e-244
2	291	91.8	506	117	HCVHN3	Hepatitis C virus cor	4,43e-221
3	285	89.9	573	116	HCUI0204	Hepatitis C virus iso	7,81e-216
4	285	89.9	573	116	HCUI0193	Hepatitis C virus iso	7,81e-216
5	285	89.9	577	69	E05083	DNA encoding a part o	7,81e-216
6	285	89.9	9400	122	HCPCGNOM	Hepatitis C China vir	7,81e-216
7	285	89.9	9410	122	HPCK1S1	Hepatitis C virus (st	7,81e-216
8	285	89.9	9410	122	HPCK1R1	Hepatitis C virus (st	7,81e-216
9	284	89.6	1613	122	HPCNS1SPC	Hepatitis C virus cor	5,84e-215
10	283	89.3	327	143	A50350	Sequence 5 from Paten	4,36e-214
11	283	89.3	346	117	HCVRIN5	Hepatitis C virus (ge	4,36e-214
12	283	89.3	573	116	HCUI0209	Hepatitis C virus iso	4,36e-214
13	283	89.3	1863	122	HPCHC32	Hepatitis C virus (HC	4,36e-214
14	283	89.3	9410	122	HPCK1R2	Hepatitis C virus (st	4,36e-214
15	283	89.3	9410	122	HPCK1S2	Hepatitis C virus (st	4,36e-214
16	283	89.3	9427	122	HPCTRNA	Hepatitis C virus str	4,36e-214
17	282	89.0	326	122	HPCHCV006	Hepatitis C virus gen	3,26e-213
18	282	89.0	326	122	HPCHCV007	Hepatitis C virus gen	3,26e-213
19	282	89.0	9535	115	D85516	Hepatitis C virus gen	3,26e-213
20	281	88.6	506	117	HCVHB1	Hepatitis C virus cor	2,43e-212
21	281	88.6	506	117	HCVHB2	Hepatitis C virus cor	2,43e-212
22	281	88.6	573	116	HCUI0225	Hepatitis C virus iso	2,43e-212
23	281	88.6	573	116	HCUI0203	Hepatitis C virus iso	2,43e-212
24	281	88.6	573	116	HCUI0189	Hepatitis C virus iso	2,43e-212
25	281	88.6	573	116	HCUI0202	Hepatitis C virus iso	2,43e-212
26	281	88.6	848	117	HCU63376	Hepatitis C virus cor	2,43e-212
27	281	88.6	906	68	E03572	cDNA sequence encodin	2,43e-212
28	281	88.6	1551	122	HPCSTRUCTB	Hepatitis C virus cor	2,43e-212
29	281	88.6	1620	122	HPCE2COR	Hepatitis C virus cor	2,43e-212
30	281	88.6	1880	122	HPC5TRJ4	Hepatitis C virus RNA	2,43e-212
31	281	88.6	2540	68	E04760	cDNA encoding a part	2,43e-212
32	281	88.6	2540	69	E04805	cDNA to 5'-terminal r	2,43e-212
33	281	88.6	2540	69	E07391	cDNA encoding chimpan	2,43e-212
34	281	88.6	4129	122	HPCPOLYPRO	Hepatitis C virus (HC	2,43e-212
35	281	88.6	9391	68	E04721	DNA encoding peptide	2,43e-212
36	281	88.6	9391	69	E04579	Gene coding Hepatitis	2,43e-212
37	281	88.6	9414	122	HCUNKNGS	Hepatitis C virus mRN	2,43e-212
38	281	88.6	9415	116	HCUI6362	Hepatitis C virus pol	2,43e-212
39	281	88.6	9440	124	S62220	polyprotein [Hepatis	2,43e-212
40	281	88.6	9448	122	HPGJ483	Hepatitis C virus gen	2,43e-212
41	281	88.6	9456	122	HPCRNA	Hepatitis C virus RNA	2,43e-212
42	280	88.3	326	122	HPCHCV010	Hepatitis C virus gen	1,82e-211
43	280	88.3	326	122	HPCHG2F	Hepatitis C virus (is	1,82e-211
44	280	88.3	326	122	HPCHCV077	Hepatitis C virus gen	1,82e-211
45	280	88.3	2433	143	A48711	Sequence 49 from Paten	1,82e-211

ALIGNMENTS

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

mpsrch nn n.a. - n.a. database search, using Smith-Waterman algorithm

```
Run on: Mon Feb 23 20:43:03 1998; MasPar time 435.69 Seconds
1066.03 Million cell updates/sec
```

Tahvil ar

```

Title:
>US-08-836-075A-1
Description:
(1-327) from US08836075A.seq
317
Perfect Score:
N.A. Sequence:
1 ATGAGCACCAATCTTAACC.....CTCGGCCAATTGGGGCCC 327
TACTCTGCTTAGATTGG.....GAGCGGGTTAACCCGGGG
Comp:

```

Scoring table: TABLE default

Match STN : Phase 0: Query 0

Searched: 430261 seqs. 710217276 bases x 2

Post-processing: Minimum Match 0%

Database: embl-new7

1:BCT 2:FUN 3:GEN 4:HTG1 5:HTG2 6:HTG3 7:HTG4 8:HUM1
9:HUM2 10:HUM3 11:INV1 12:INV2 13:ORG 14:MAM 15:VRT
16:PLN 17:PRO1 18:PRO2 19:ROD 20:SYN 21:UNC 22:VIR
genbanka[0]

Database:

23:BC11 24:BC2 25:BC3 26:BC4 27:BC5 28:BC6 29:BC7
 30:BC8 31:BC9 32:BC10 33:BC11 34:BC12 35:BC13
 36:GEN1 37:GEN2 38:GEN3 39:GEN4 40:GEN5 41:GEN6 42:HTG1
 43:HTG2 44:HTG3 45:HTG4 46:HTG5 47:INV1 48:INV2 49:INV3
 50:INV4 51:INV5 52:INV6 53:INV7 54:INV8 55:INV9 56:INV10
 57:INV11 58:INV12 59:MAM1 60:MAM2 61:MAM3 62:VRT1
 63:VRT2 64:VRT3 65:VRT4 66:PAT1 67:PAT2 68:PAT3 69:PAT4
 70:PAT5 71:PAT6 72:PAT7 73:PLN1 74:PLN2 75:PLN3
 77:PLN4 78:PLN5 79:PLN6 80:PLN7 81:PLN8 82:PLN9 83:PLN10
 84:PLN11 85:PLN12 86:PR11 87:PR12 88:PR13 89:PR14
 90:PR15 91:PR16 92:PR17 93:PR18 94:PR19 95:PR110
 96:PR111 97:PR112 98:PR113 99:PR114 100:PR115 101:PR116
 102:PR117 103:ROD1 104:ROD2 105:ROD3 106:ROD4 107:ROD5
 108:ROD6 109:ROD7 110:ROD8 111:ROD9 112:STR 113:SYN
 114:UNA
 genbankb101
 115:VRL1 116:VRL2 117:VRL3 118:VRL4 119:VRL5 120:VRL6
 121:VRL7 122:VRL8 123:VRL9 124:VRL10 125:VRL11
 genbank-nc7
 126:BC1 127:GEN 128:HTG1 129:HTG2 130:INV 131:MAM

Database:

Database:

hbank-new / 126:BCT 127:GEN 128:HTG1 129:HTG2 130:INV 131:MAM

RESULT 1
ID A50346 standard; DNA; UNC; 327 BP.
AC A50346;
NI e1039185
DT 07-MAR-1997 (Rel. 51, Created)
DE 07-MAR-1997 (Rel. 51, Last updated, Version 1)
DE Sequence 1 from Patent W09613590.
KW
OS unidentified
OC unclassified.
RN [1]
RP 1-327
RA Maertens G., Stuyver L.;
RT "NEW SEQUENCES OF HEPATITIS C VIRUS GENOTYPES AND THEIR USE
AS PROPHYLACTIC, THERAPEUTIC AND DIAGNOSTIC AGENTS";
RL Patent number W09613590-A/1, 09-MAY-1996.
RL INNOGENETICS NV (BE).
CC Other publication AU 3844095 960523
FH Key Location/Qualifiers
FH source 1..327
FH /organism="unidentified"
SQ Sequence 327 BP; 56 A; 102 C; 104 G; 51 T; 14 other;
Query Match 100.0%; Score 317; DB 143; Length 327;
Best Local Similarity 100.0%; Pred. No. 7.68e-244;
Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 atgagcagcaatcctaaacctcaaaagaaacaaacgtaacacacacccgcgcctcaak 60
Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACAGGTAAACACCAACCGCGCCTCAK 60
Db 61 gsgtgnnnnnnccgggtggcggtcagatcgttggtgagttacacctgttgcgcgcagg 120
Qy 61 GSGGTNNNNNNCCGGGTGGCGGTCAAGATCGTGTGGAGCTTACCTGTGGCGGCAGG 120
Db 121 ggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 180
Qy 121 GGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
Db 181 aggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 240
Qy 181 AGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
Db 241 taccctggccctctatggcaatgaggggtgcgggtggcggtggcggtggcggtggcggt 300
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGGGTGCGGTTGGCGGTGGCGGTGGCGGTGG 300
Db 301 cgcggctcgc 327
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 2
LOCUS HCV/HN3 506-bp DNA VRL 29-APR-1994
DEFINITION Hepatitis C virus core gene, HN3 isolate.
ACCESSION X76410
NID 6434947
KEYWORDS core; core protein.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.

REFERENCE 1 (bases 1 to 506)
AUTHORS Qu, D.
TITLE Direct Submission
JOURNAL SUBMITTED (22-NOV-1993) to the EMBL/GenBank/DBJ databases. D. Qu,
INSERM Unite 271, U. de Recherche sur les Hepatites, le Sida et
les Retrovirus Humains, 151 cours Albert Thomas, F 69424 Lyon Cedex
03, FRANCE
REFERENCE 2 (bases 1 to 506)
AUTHORS Qu, D., Hantz, O., Gouy, M., Vitvitski, L., Li, J.S., Berby, F.,
Tong, S.P. and Trepo, C.
TITLE Heterogeneity of hepatitis C virus genotypes in France
JOURNAL J. Gen. Virol. 75 (Pt 5), 1063-1070 (1994)
MEDLINE 94231157
FEATURES Location/Qualifiers
source 1..506
/organism="Hepatitis C virus"
/isolate="HN3"
misc_feature 29..506
/notes="region corresponding to the core protein of the
polyprotein of HCV"
BASE COUNT 86 a 166 c 166 g 88 t
ORIGIN
Query Match 91.8%; Score 291; DB 117; Length 506;
Best Local Similarity 91.7%; Pred. No. 4.45e-221;
Matches 300; Conservative 4; Mismatches 23; Indels 0; Gaps 0;
Db 29 atgagcagcaatcctaaacctcaaaagaaacaaacgtaacacacccgcgcgcagcag 88
Qy 1 ATGAGCAGCAATCCTTAACCTCAAGAAACCAACAGGTAAACACCAACCGCGCCTCAK 60
Db 89 gacgtcgaagtcctccggcggtggtcagatcgttggtgagttacacctgttgcgcgcagg 148
Qy 61 GSGGTNNNNNNCCGGGTGGCGGTCAAGATCGTGTGGAGCTTACCTGTGGCGGCAGG 120
Db 149 ggcacagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 208
Qy 121 GGCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 180
Db 209 aggcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcagcag 268
Qy 181 AGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 240
Db 269 taccctggccctctatggcaatgaggggtgcgggtggcggtggcggtggcggtggcggt 328
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGGGTGCGGTTGGCGGTGGCGGTGGCGGTGG 300
Db 329 cgtggtctcgc 355
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 3
LOCUS HCV10204 573 bp RNA VRL 26-AUG-1994
DEFINITION Hepatitis C virus isolate P10 core protein gene.
ACCESSION U10204
NID g532392
KEYWORDS Hepatitis C virus.
SOURCE Hepatitis C virus.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (g1-ss)
AUTHORS Bukh, J., Purcell, R.H. and Miller, R.H.

23
14
9
318
327 - 977b

Feb 23 20:47

/home/fuller/feb1098/US-08-836-075A-1.rge

5

```
TITLE      Sequence analysis of the core gene of 14 hepatitis C virus
            genotypes
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 91, 8239-8243 (1994)
MEDLINE    94336721
REFERENCE  2 (bases 1 to 573)
AUTHORS    Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE      Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE    92279243
REFERENCE  3 (bases 1 to 573)
AUTHORS    Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE      At least 12 genotypes of hepatitis C virus predicted by sequence
            analysis of the putative E1 gene of isolates collected worldwide
JOURNAL    Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8234-8238 (1993)
MEDLINE    93376778
REFERENCE  4 (bases 1 to 573)
AUTHORS    Bukh, J.
TITLE      Direct Submission
JOURNAL    Submitted (31-MAY-1994) Bukh J., National Institutes of Health,
            Hepatitis Viruses Section, 9000 Rockville Pike, Bethesda, MD 20892
            USA
```

FEATURES

```
source      Location/Qualifiers
            1..573
            /organism="Hepatitis C virus"
            /isolate="p10"
            /specific_host="Homo sapiens"
            1..573
            /gene="C"
            /standard_name="core"
            /function="nucleocapsid protein"
            /evidence=experimental
            /product="core protein"
            1..573
            /gene="C"
            1..>573
            /gene="C"
            /codon_start=1
            /product="polyprotein"
            /db_xref="PID:g532393"
            /translation="MSTNPKPQRTKNTNRRPQDVKFGGGQIVGGVYLLPRGPRL
            GVRATKTSERSQPRGRQIPKARRPEGRAWAQGYPNPLYCNEGLWAGWLLSPRG
            SRP SWGFTDPRRRNLGKVIDTLTCGFADLMGYIPLVGAPLGGARALAHGVRVLED
            GVNATGNLFGGSFIFLLALLSCITIPASA"

mat_peptide 96 a 185 c 179 g 113 t
```

```
BASE COUNT 96 a 185 c 179 g 113 t
ORIGIN
Query Match      89.9%; Score 285; DB 113; Length 573;
Best Local Similarity 90.8%; Pred. No. 7.91e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Dbb 1 atgacgacgaatcctaaacctcaagaacaaacaaactaaccacacgcgcgcacag 60
|||||
Qy 1 ATGACGACCAATCCTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||

Dbb 61 gacgtcaagtcccgggcggtggtcagatcgttggtgagtttaacctgtgcgcgcagg 120
|||||
Qy 61 GCGGTNNNNNNCCGGTGGCGGTGACATCGTGTGCGAGTTTACCTGTTCGCGCAGG 120
|||||

Dbb 121 ggccccaggttggtgtggtgcgagactaggaagacttccgagcggtcgcaacctcgtgga 180
|||||
Qy 121 GCGCCCAAGNNNGGTGTGTCGGCGCACTAGGAAGACTTCGAGCGGTCAACACCTCGTGGC 180
|||||

Dbb 181 aggcgcacaacctatccccaaaggttcgcccgcagggcaggcgacctgggctcagccggg 240
|||||
```

Feb 23 20:47 /home/fuller/feb1098/US-08-836-075A-1.rge

6

```
Qy 181 AGGCACGACGCTATCCCAAGGCTCGYCGGTCGAGGGCAGGTCTTGGGCTCAGCCCGGG 240

Dbb 241 taccttgcccctctatggaatgagggcttgggtggcagatggctctctcacc 300
|||||
Qy 241 TATCCTTGCGCCCTCTATGGAATGAGGCTCGCGGTGGCGGGTGGCTCTCTCCCC 300
|||||

Dbb 301 cgtggtctcgccctagttggggcccc 327
|||||
Qy 301 CGCGGCTCTCGGCGCAATTGGGGCCCC 327
|||||
```

```
RESULT 4
LOCUS     HCU10193      573 bp      RNA          VRL      26-AUG-1994
DEFINITION Hepatitis C virus isolate DK1 core protein gene.
ACCESSION U10193
NID       g532370
KEYWORDS  .
SOURCE    Hepatitis C virus.
ORGANISM  Hepatitis C virus
           Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
           Hepatitis C-like viruses.
REFERENCE 1 (sites)
AUTHORS   Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE     Sequence analysis of the core gene of 14 hepatitis C virus
           genotypes
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 91, 8239-8243 (1994)
MEDLINE   94336721
REFERENCE 2 (bases 1 to 573)
AUTHORS   Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE     Sequence analysis of the 5' noncoding region of hepatitis C virus
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)
MEDLINE   92279243
REFERENCE 3 (bases 1 to 573)
AUTHORS   Bukh, J., Purcell, R.H. and Miller, R.H.
TITLE     At least 12 genotypes of hepatitis C virus predicted by sequence
           analysis of the putative E1 gene of isolates collected worldwide
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8234-8238 (1993)
MEDLINE   93376778
AUTHORS   Bukh, J.
TITLE     Direct Submission
JOURNAL   Submitted (31-MAY-1994) Bukh J., National Institutes of Health,
           Hepatitis Viruses Section, 9000 Rockville Pike, Bethesda, MD 20892
           USA
```

```
FEATURES      Location/Qualifiers
source        1..573
              /organism="Hepatitis C virus"
              /isolate="DK1"
              /specific_host="Homo sapiens"
              1..573
              /gene="C"
              /standard_name="core"
              /function="nucleocapsid protein"
              /evidence=experimental
              /product="core protein"
              1..573
              /gene="C"
              1..>573
              /gene="C"
              /codon_start=1
              /product="polyprotein"
              /db_xref="PID:g532371"
              /translation="MSTNPKPQRTKNTNRRPQDVKFGGGQIVGGVYLLPRGPRL
              GVRATKTSERSQPRGRQIPKARRPEGRAWAQGYPNPLYCNEGLWAGWLLSPRG
```

SRP SNGNDP RRRSRLKRV IDTLTCGFADLMGY IPIVGA PLGGAARALAHGVRLIED
GNYATGNLP GCSFISFILALLSCLITIPASA*
BASE COUNT 98 a 185 c 177 g 113 t

Query Match 89.9%; Score 285; DB 116; Length 573;
Best Local Similarity 90.8%; Pred. No. 7.81e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 1 atgagcagcaatcctaaacctcaaaagaaacaaacgtacacacccgcccacag 60
|||||
Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 61 gacgtcaagttcccggtggtgcagatcgttggtgagtttacctgttgcgcgcag 120
|||||
Qy 61 GCGTNNNNNNCCGGTGGCGGTGAGATCGTGTGGTGGAGTTTACCTGTGCGCGCAGG 120
|||||
Db 121 gcccaggttggtgtgcgcgactaggaagacttcgagcgttcgcaacctcgtgga 180
|||||
Qy 121 GCGCCAGGNNGGGTGCGCGGACTAGGAAGACTTCCGAGCGGTCAACAACCTCGTGGC 180
|||||
Db 181 agcgcaaacctatcccaagctcgcgcgagggcagggcagggcctgggctcaagccgg 240
|||||
Qy 181 AGCGCAGACGCTATCCCAAGGCTCGCGYCCGAGGGCAGGCTCGGCTCAGCCCGG 240
|||||
Db 241 taccctggccctctatggcaatgagggcagtggtggcagagtggtcctctgtcaccc 300
|||||
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGCGGGGNTGGCTCTGTCCCC 300
|||||
Db 301 cggcgctcggcctagttggggcccc 327
|||||
Qy 301 CGCGGCTCGGCCCAATGGGGCCCC 327
|||||

RESULT 5
LOCUS E05083 577 bp RNA PAT 26-NOV-1996
DEFINITION DNA encoding a part of antigen of type non-A non-B hepatitis virus.
ACCESSION E05083
NID g2173277
KEYWORDS JP 1993176774-A/20.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 577)
AUTHORS Kunitada, S., Muneo, T., Rieko, F., Kaoru, S., Masako, C., Hajime, H., Naofumi, H., Akimasa, O., Noriyoshi, M., Takashi, K., Takeshi, T., Haruhide, K., Kazumasa, Y., and Ken, O.
TITLE HEPATITIS NON-A-NON-B ANTIGEN FRAGMENT, ITS PRODUCTION AND REAGENT FOR HEPATITIS NON-A-NON-B DIAGNOSIS USING THE SAME
JOURNAL Patent: JP 1993176774-A 20-JUL-1993;
SHIMOTOONO KUNITADA, GREEN CROSS CORP:THE
COMMENT OS type non-A non-B hepatitis virus
PN JP 1993176774-A/20
PD 20-JUL-1993
PF 18-DEC-1991 JP 1991354708
PR 18-DEC-1990 JP 90P 412020
PI SHIMOTOONO KUNITADA, TSUJIKAWA MUNE0, FURUTA RIEKO, PI SAITO KAORU,
PI CHUGENJI MASAKO, HORII HAJIME, HAYASUKE NAOFUMI, OMIZU AKIMASA, PI MIYANO NORIYOSHI, KOBAYASHI TAKASHI, TAKAHASHI TAKESHI, PI KAWABE HARUHIDE.
PI YOKOYAMA KAZUMASA, OKABAYASHI KEN
PC C12N15/51, C07K13/00, C12P21/02, G01N33/53, G01N33/569, G01N33/576,
PC (C12P21/02,

PC C12R1:19);
CC strandedness: Double;
CC topology: Linear;
CC hypothetical: No;
CC anti-sense: No;
FEATURES Location/Qualifiers
source 1..577
/organism="unidentified"
BASE COUNT 103 a 183 c 179 g 112 t
ORIGIN

Query Match 89.9%; Score 285; DB 69; Length 577;
Best Local Similarity 90.8%; Pred. No. 7.81e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 39 atgagcagcaatcctaaacctcaaaagaaacaaacgtacacacccgcccacag 98
|||||
Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60
|||||
Db 99 gacgtcaagttcccggtggtgcagatcgttggtggagtttaacctgtgcgcgcag 158
|||||
Qy 61 GCGTNNNNNNCCGGTGGCGGTGAGATCGTGTGGTGGAGTTTACCTGTGCGCGCAGG 120
|||||
Db 159 gcccaggttggtgtgcgcgactaggaagacttcgagcgttcgcaacctcgtgga 218
|||||
Qy 121 GCGCCAGGNNGGGTGCGCGGACTAGGAAGACTTCCGAGCGGTCAACAACCTCGTGGC 180
|||||
Db 219 agcgcaaacctatcccaagctcgcgcgagggcagggcagggcctgggctcagccggt 278
|||||
Qy 181 AGCGCAGACGCTATCCCAAGGCTCGCGYCCGAGGGCAGGCTCGGCTCAGCCCGG 240
|||||
Db 279 taccctggccctctatggcaacgagggctcgggagtggtgggagtggtcctctgtcaccc 338
|||||
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGCGGGNTGGCTCTGTCCCC 300
|||||
Db 339 cggcgctcggcctagttggggcccc 365
|||||
Qy 301 CGCGGCTCGGCCCAATGGGGCCCC 327
|||||

RESULT 6
LOCUS HPCCGENOM 9400 bp ss-RNA VRL 28-SEP-1992
DEFINITION Hepatitis C China virus complete genome.
ACCESSION L02836
NID g329739
KEYWORDS complete genome; viral genome.
SOURCE Hepatitis C virus (individual isolate China, strain HeBei) (Library: NC1-18) cDNA to genomic RNA.
ORGANISM Hepatitis C virus
VIRUSES; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (bases 1 to 9400)
AUTHORS Bi, S.-L., Bai, X.-H., Margolis, H.S. and Liu, C.-B.
TITLE Complete hepatitis C virus genome cDNA isolated from mainland of China
JOURNAL Unpublished (1992)
FEATURES Location/Qualifiers
source 1..9400
/organism="Hepatitis C virus"
/isolate="China"
/strain="HeBei"
/tissue lib="NC1-18"
355...9384
mRNA 1920 a 2796 c 2650 g 2034 t
BASE COUNT

FNMAYTKIKLTP IPAASQLDLSNFWVAGYSGDVYHSLSRPRMFMLCILLISVGV

```
mat_peptide 330..902
mat_peptide 903..1478
mat_peptide 1479..2756
misc_feature 1479..1559
mat_peptide 2757..3407
mat_peptide 3408..5300
mat_peptide 5301..5462
mat_peptide 5463..6245
mat_peptide 6246..7586
misc_feature 6954..7073
mat_peptide 7587..9359
BASE COUNT 1904 a 2846 c 2691 g 1969 t
ORIGIN
```

```
Query Match 89.9%; Score 285; DB 122; Length 9410;
Best Local Similarity 90.8%; Pred. No. 7,81e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 330 atgagcagcaatcctaaccctcaaaagaaacaaacgttaacacacacccgcgcacag 389
|||||
Qy 1 ATGAGCAGCAATCCTAAACCTCAAGAAACCAACACGTTAACCAACCGCCGCTCAK 60
|||||

Db 390 gacgtcaagtcccgccggtggtcagatcgttggtgagttacctgtgcccgcagg 449
|:|:|
Qy 61 GCGCTNNNNNNCCGGCTGCGCGTCAGATCGTGTGCGAGTTACTGTGTCGCGCAGG 120
|||||

Db 450 ggcacaggttggtgtgcccgcagactaggaaactccgagcgtgcgaacctcgtgga 509
|||||
Qy 121 GCGCCACAGNNGGCTGCGCGCGACTAGGAGACTTCCGACGCGTCACACCTCGTGC 180
|||||

Db 510 aggcgaacacctatccccagagctcgccgcgagggcaggcggcctgggctcagccggg 569
|||||
Qy 181 AGCGGACAGCCTATCCCCAGGCTCGYCGGTCGAGGGGCGAGCTCCTGGCGTCAGCGCGG 240
|||||

Db 570 taccctggccctctatgcgaatgagggcttaggtgggcaggtgacctcctgcaacc 629
|||
Qy 241 TATCCTTGCGCCCTCTATGCGCAATGAGGCTCGCGGTGCGGCGGNGTGGCTGTGCCCC 300
|||||

Db 630 cgcgcgtcccgccctagttggggccccc 656
|||||
Qy 301 CGCGGCTCGGCCCAATTGGGGGCCCC 327
|||||
```

```
RESULT 8
LOCUS HPCK1R1 9410 bp as-RNA VRL 02-FEB-1997
DEFINITION Hepatitis C virus (strain HCV-1b, clone HCV-K1-R1), complete genome
sequence.
ACCESSION D50480
NID g1030706
KEYWORDS polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
NS5B; envelope protein; non-structural protein;
```

```
interferon-sensitive; interferon-resistant; IFN-sensitive;
IFN-resistant; ISDR; interferon sensitivity determining region;
HVR; hypervariable region.
SOURCE Hepatitis C virus (strain:HCV-1b) cDNA to genomic RNA,
clone:HCV-K1-R1.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
REFERENCE 1 (sites)
AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
TITLE Comparison of full-length sequences of interferon-sensitive and
resistant hepatitis C virus 1b. Sensitivity to interferon is
conferred by amino acid substitutions in the NS5A region
J. Clin. Invest. 96 (1), 224-230 (1995)
JOURNAL 95340824
MEDLINE
REFERENCE 2 (bases 1 to 9410)
AUTHORS Enomoto,N.
JOURNAL Unpublished (1995)
REFERENCE 3 (bases 1 to 9410)
AUTHORS Enomoto,N.
TITLE Direct Submission
JOURNAL Submitted (08-MAY-1995) to the DDBJ/EMBL/GenBank databases.
Hobuyuki Enomoto, Tokyo Medical and Dental University, Second
Department of Internal Medicine; 1-5-45 Yushima, Bunkyo-ku, Tokyo
113, Japan (E-mail:PXN04522@niftyserve.or.jp,
Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)
```

```
FEATURES
source
1..9410
/organism= "Hepatitis C virus"
/notes= "IFN-resistant HCV from patient 1"
/strain= "HCV-1b"
/clones= "HCV-K1-R1"
330..9362
/codon_start=1
/products= "polyprotein"
/db_xref= "PID:d1009706"
/db_xref= "PID:g1814085"
/tranlation= "MSTNPKQRKTKNTNRPDQVKPGGGQIVGGYVLLPRGPRLL
GYRATKTSERQPRRRQPIFKARRPREGMAQPGTPWFLYNGEGLGAGHLLSPRG
SRPSGPTDPRRRSRNLGRVIDTLTCGFADLMGYTPIVCGAPLCGAARALAHGVRIED
GVNATGNLPCCFSIFLLALLSLCITIPASAVEVCNAGLHVHTDCSNSIIVYEAQD
MIMTPGCVPCVRENNSSRCWALTPTLAARNASIPFTTIRHVDLLVGAACCSAMY
VGDLCGSVFVLSQFLTFSPPRYETVDONCSIYPGHVSGHRMAMMMNWSPTTALVY
SOLLRIPOAVDMVAGAHMGLAGLAYYSMVGNWAKVLIIVMLLFAGVGNVTTTGAQ
GRTVSRLLTSITPGASQNTQLINTNGSHINRTALNCNDSLOTGFIALLEYARRENS
GQPRLASVRSIDKEAQMGPIITYAEHSDQRPYQWRYAPRPGCTPIPAQVCGPVYC
FTTPSPVGTGTDRECAPTYNMGANETDVLANTTPQGNMFCCTMNGTGFTKTCGG
PPCNIIGCAGNNTLCTDTCFRKHEATYAKCGSGPMLTPRCIVDYPVRLMHPYCTVNF
SIFKRYMYGGVGHRLTAACNWRGERCDLEDORSELSPLLSTTEMQILPCSTFTLL
PALSTGLIHQNTVDVQVLYGVGSVAVSFAIKMEYVLLFLLLADARVACALWMLLL
VAQAAALENIIVLNNAASLAGVHGLSFLVEFCAMWYIKGKIVPGCAAYLVGWFLLL
LLIALPPRAYAMDREMAASCGGVAIGLALLTSPYKVF LAKLILWLIQYILTRAEAQ
LQWNIPIINVGRGDAIILLTCAPHEIFDITKIVLAILGCLMVLQAGTAKYVPDFVR
AQCIVRECTLVKRYGGGVQVQMAFMKLAELKGTSVYDHLTPLRDWAHTGLRDLAVAVE
PVFSDMEIKIITWGGDTTACGDIIINGLPVSARRGRELILGPADSLECGQGRLLAPIT
AYSQQTREGLFGCIITISITGRDNQVEGEVQVVSSTATQSF LATICVNGVCWTVYHGASK
TLAGKGPITQMYTNVDQLVGMAPPARSLTPTCTCGSDILVLRHADVLPVRRRG
DSRGSLLSRPISYIKCSGCGLLCPSCHVVGIFRAAVCTRGVAKAVDFVPVSMETT
MRSVFTDNSSPPVPQFTQVAHLHAPTGSKSTKVPAAQAQGYKVLVLPVSTATIL
GFGAYMSKAGHVDNIRITGVRTITTTGAPITITYSTYKGLADGCGSGGAYDIMCDECHS
IDSTSILGCTVLDQAEACARILVLAATATPPTCVTPHPNIEETALSNGEIPFYK
AIPITETKGRHLIFCHSKKKDELANLKSGLGLNNAVYYRGLDVSVIPASGVVVA
TDALMTGTGDFSDVDCNTCVTQTVDFSLDPTFTTETTTVPQDAVSRSQRGRTGRG
```

CDS

	1994	a	2850	c	2693	g	1973	t
mat_peptide	330	..	902					
mat_peptide	903	..	1478					
mat_peptide	1479	..	2756					
misc_feature	1479	..	1559					
mat_peptide	2757	..	3407					
mat_peptide	3408	..	5300					
mat_peptide	5301	..	5462					
mat_peptide	5463	..	6245					
mat_peptide	6246	..	7586					
misc_feature	6954	..	7073					
mat_peptide	7587	..	9359					

BASE COUNT 1994 a 2850 c 2693 g 1973 t

ORIGIN

Query Match 89.9%; Score 285; DB 122; Length 9410;
Best Local Similarity 90.8%; Pred. No. 7.81e-216;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 330 atqacacgaatcctaaacctcaaaqaaaaaaccaaacqtaacaccacccqcccccacq 389

1 ATCAGCAGCAATCCTAAACCTCAAGAGAAAACCAACGTATACACCAAGCGCGCCCTCAK

0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

Db	450	ggccccaggttggtgtgcgcgcagactccgagcagctcgcaacctcgtgga	509
Qy	121	ggccccagcagnnngggtgtgcgcgcgacttaggaagacttccgagcgggtcacaacctcgtggc	180
Db	510	aggcgacaacctatcccacaagctcgccgcgcgcgagggcaggctggctcagcccg	569
Qy	181	aggcgacagcctatccccaaggctgcgcgcgcgagggcaggtctcgggtcagcccg	240
Db	570	taccttgcccctctatggcaatgaggcttaggtggcgaggatggctcgtgcacc	629
Qy	241	tatccttggccccctctatggcaatgagggttgcgggtggcggggtgctcctgtcccc	300
Db	630	cgcggctccgcgcctagtgtggggccc	656
Qy	301	cgccggctcgcggccaaattggggccc	327

RESULT	9	
LOCUS	HPNCNS1SPC	1613 bp ss-RNA
DEFINITION	Hepatitis C virus core protein	VRL 03-DEC-1992
ACCESSION	M74806	(NS1) gene.
NID	g329784	
KEYWORDS	NS1 protein; core protein; envelope-associated protein.	
SOURCE	Hepatitis C virus (individual isolate SPI)	cDNA to genomic RNA.
ORGANISM	Hepatitis C virus	
	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;	
	Hepatitis C-like viruses.	

REFERENCE	1 (bases 1 to 1613)
AUTHORS	Chae, T.-A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G. and Urdea, M. S.
TITLE	At least five related, but distinct, hepatitis C viral genotypes exist
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 89, 7144-7148 (1992)
MEDLINE	92357788
FEATURES	Location/Qualifiers

```

1.1613
/organism="Hepatitis C virus"
/isolate="SPI*"

290 a 501 c 479 g 343 t

BASE COUNT
ORIGIN

```

Query Match 89.6%; Score 284; DB 122; Length 1613;
Best Local Similarity 90.8%; Pred. No. 5.84e-215;
Matches 295; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 244 atgagcaggaatcctaacctcaagaaaaaaccaaacgtaacaccaaccgcgcacag 303
|||||

Db 304 gacgtcaagtcccgggcggtggtcagatcgttggtggagtttacctgttgccgcgcagg 363

Qy 61 GGSCTNNNNNNCCGGGTGGCGGTACATCGTTGGTGGAGTTTACCTGTTGCCCGGCAGG 120

Qy 121 GGGCCAGGNNGGGTGTGGCGCGACTAGGAAGACTTCCGAGCGGTACACCTCGTGGC 180

[illegible]

Db 484 t a t c c t t g g c c c c t c t a t g g c a a t g a g g g c c t g g g g t g g g c a g g a t g g c t c c t g t c a c c c 543

QY Z4T TATCCITGGCCCCCTCATAAGGCAATGAGGGGCATCGGGGGTGGGTTGGGTTCCTGTCCTG

RRGYRVVTPGRSPGMFSSVCECYDAGCAWELTPAETSVRLRAYLNTLPCLPFCQ
 DRIIEFEGVFTGLTHIDAHFSQTKAGENLPYLVAYQVACARAQAQPPSPDMWMKC
 LTRUKPLTHGPTLLXRLGAVONDITLTHPTIKKYIMTCSADLEAVTSTWLVGGVLA
 AUTAYCLTUGSSVIVGRIILSKRAVAPIDREALYOEDEMEECASHLPIYEOQMOLAE
 GEPKOKALQLQATNAQEAAPVWESKWARLEAFWAKHMFNFISCIQYLAGLSTLPGN
 PAIASLMAFTASTISPTLSTQTLNLTIGKVAQAAPPRAVAFVCGAGIACAAGVISJ
 GLGKVLVDILAGYAGVAGALVAFKWSGEMPSTEDLVNLLPALISGALVGVWCVA
 LTRHVRGPEGAVQNMNRLAPFASRGNHNSPHTYWPESDAAARVOTLASLTJQLRL
 RLHWMINEDCSTPCGSGSNDVDMOICTVLDFKMLQSKLLPLRGVPPFSFCORGVR
 GWRGDCILMQTCTCGGAQITGVHNGCSMRIVGPKTCNTWHTGFTFINAYTGPCTPSP
 APNYSRALMARAAVEYVGLNOYGVHVTGMNCTDNVCKQVPPPEFTTEVDVQVRL
 NAAACGPLREEVFTYVGLNOYLVGSGDLPCEPDVTLTSMULTPDSHTIGEAARKRL
 ARGSPSSCLASSASQLSAPSKACTTTHDSDPADLEANLLHQRQMGONIIRVSESEN
 KVIITDSDPTLRAAEDEREVSPTAEIRIKSKFSPALPITWAPRDYNNPILLESXKDQPT
 VPPVYHCGPLPPAKAFIPPPRRKPTLITESTVSSALAEIATKTCGSGSSADRCPT
 ATAPDQASNUGDAGSDVGSYSSMPLLEGEPGDPDSDGCSWSTVSEAEADVVCSSMS
 YTWGALITPCAAESELPIINALSNLILRHNNHYATTSRASQORQKRVFTDRLQVID
 DHYRDLVKMCAKAKASTVKAKLLLSVEACRLTPHSARSKEFYCAKDVRNLSGAVNHJ
 RSWKNOJLEDTEPTIDTTIMAKNEVFCVQPEKGGKKEPARIIVPDLGVVRCEKALYDV
 WYSLPQAGSSGSGFYDSPQORVEVLWNAKSKKCPMGFAVYDTRCFDSTVTESDIRV
 EESTYQCDLAPFAEQARIRSLVRIITGCPRLTNKQKQNGCYRRACRSGVILTSCONT
 TCYKLSAASAKRAKLODCTMLVCGDILVVICESATQEODASILRVFTFAMRTYSNPPG
 DPPQEPYDLELITSCSSNVAHVASAKRVLVITRDPTTTLARAAWETAKSTPVNSML
 KNIIMFAPTLJWYRMIINLTHVHFSJLILAQEQLAKDLQCITYGACYSIEPDIQJLQRLH
 GLSAFSILHYSYSPGEINNVASCLKLGVPPLRTWRHRRARSVRAKLLSOGGRANTCGRYL
 FNWAVKTKLUTPPIPAASQJLDJUNWFFVAGYSVCGDVYHISLSRARPWFMLCJLLLSVGV
 GYLLPLNR*

```
mat_peptide      330..902
                  /product="core protein"
mat_peptide      903..1478
                  /product="EI"
mat_peptide      1479..2756
                  /product="E2"
misc_feature      1479..1559
                  /note="hypervariable region (HVR)"
```

```
mat_peptide
2757..3407
/product="NS2"

mat_peptide
3408..5300
/product="NS3"

mat_peptide
5301..5462
```

```
mat_peptide
5463..6245
/product="NS4B"
6246..7586

mat_peptide
```

```
misc_feature
6954..7073
/note="interferon sensitivity determining region (ISDR)"
```

BASE COUNT	1894 a	2850 c	2693 g	1973 t
/product="NSB"				

Query Match	89.9%;	Score 285;	DB 122;	Length 9410;
	0.022	7.0	0.0	0.0

Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Db 330 atgaacgaatcctaaacctcaaaqaaaaaaccaaacqtaacaccaccccccacag 389

07 1 ATGACCAATCTTAACCTCAACAGCAAAACCAACCTAACCAACCGCCCTCAK 60

[illegible]

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
84

NID	g532402	
KEYWORDS		
SOURCE	Hepatitis C virus.	
ORGANISM	Hepatitis C virus	
	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.	
REFERENCE	1 (sites)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	Sequence analysis of the core gene of 14 hepatitis C virus genotypes	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 91, 8239-8243 (1994)	
MEDLINE	94336721	
REFERENCE	2 (bases 1 to 573)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	Sequence analysis of the 5' noncoding region of hepatitis C virus	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 89 (11), 4942-4946 (1992)	
MEDLINE	92279243	
REFERENCE	3 (bases 1 to 573)	
AUTHORS	Bukh,J., Purcell,R.H. and Miller,R.H.	
TITLE	At least 12 genotypes of hepatitis C virus predicted by sequence analysis of the putative E1 gene of isolates collected worldwide	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 90 (17), 8234-8238 (1993)	
MEDLINE	93376778	
REFERENCE	4 (bases 1 to 573)	
AUTHORS	Bukh,J.	
TITLE	Direct Submission	
JOURNAL	Submitted (31-MAY-1994) Bukh J., National Institutes of Health, Hepatitis Viruses Section, 9000 Rockville Pike, Bethesda, MD 20892, USA	
FEATURES		
source	Location/Qualifiers	
	1..573	
	/organism="Hepatitis C virus"	
	/isolate="S45"	
	/specific_host="Homo sapiens"	
	1..573	
	/gene="C"	
	/standard_name="core"	
	/function="nucleocapsid protein"	
	/evidence=experimental	
	/product="core protein"	
	1..573	
gene	/gene="C"	
CDS	1..573	
	/gene="C"	
	/codon_start=1	
	/product="polyprotein"	
	/db_xref="PID:g532403"	
	/translation="MSTNPKPQRTKNNRPQDVKFPGGQIVGGVILPFRPRGLR GVRAKRTSERSQPRGRQIPKARRECEAWAQPCHPAPLYCNEGICWAGWLLSPRG SRP SWGFTPPRRRRNLGKVIDTLCGFADLMGYIPIVGAP LGGAARAALAHGVRIED GVNATGNLPFGCSFIFLLALLSGLTITPASA"	
BASE COUNT	95 a 189 c 177 g 112 t	
ORIGIN		
	Query Match 89.3%; Score 283; DB 116; Length 573;	
	Best Local Similarity 90.5%; Pred. No. 4,36e-214;	
	Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;	
Db	1 atagcagcagatcctaactcaagacacaaacaaacgtcaacacaaacgcgcgccacag 60	
Qy	1 ATGCGACGAATCCTAAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60	
Db	61 gacgtcaagttcccggtggcggtcagatcgttggtagtttacctgtgcgcgcagg 120	

Qy	61	GSGETNNNNNCCGGGTGCGGTCAGATCCTTGGTGGACTTTACTGTTGCGCGCGAGG	120
Db	121	ggcccaagtgaggtgltgcgcgcgactaggaagactccgagcggtcacaaactcgtgga	180
Qy	121	ggccccaggnnggggtgscgcgcgactaggagagacttcccgagcggtcacaaactcgtggc	180
Db	181	cggcgaaacatcatcccaaggctcgccgcgcgagggcagggcgagggcctgggcccagccggg	240
Qy	181	agggcagagcgtatccccaaaggtctgcyggyccgagggcagggcttccgtgggctcagccggg	240
Db	241	cactcttgccccctctatggcaatgagggcttggggtggcgaggatggctcctgtcaacc	300
Qy	241	tatctcttgccccctctatggcaatgagggcttggggtggcgaggatggctcctgtccccc	300
Db	301	cgtggctcccgccctagttggggcccc	327
Qy	301	cgccgctctggcccaattggggcccc	327
RESULT	13		
LOCUS	HPHC02	1863 bp ss-RNA	VRL 05-AUG-1994
DEFINITION	Hepatitis C virus (HCV) genomic RNA for structural protein, partial sequence.		
ACCESSION	D10074		
NID	g221588		
KEYWORDS	structural protein.		
SOURCE	Hepatitis C virus (isolate:HC-J2) cDNA to genomic RNA.		
ORGANISM	Hepatitis C virus		
REFERENCE	Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae; Hepatitis C-like viruses.		
AUTHORS	Okamoto,H., Kurai,K., Okada,S., Yamamoto,K., Lizuka,H., Tanaka,T., Fukuda,S., Tsuda,F. and Mishiro,S.		
TITLE	Full-length sequence of a hepatitis C virus genome having poor homology to reported isolates: comparative study of four distinct genotypes		
JOURNAL	Virology 188 (1), 331-341 (1992)		
MEDLINE	92230232		
REFERENCE	2 (bases 1 to 1863)		
AUTHORS	Okamoto,H.		
JOURNAL	Unpublished (1991)		
COMMENT	Data kindly submitted in computer readable form by: Hiroaki Okamoto		
	Immunology Division		
	Jichi Medical School		
	Kawachi-gun		
	Tochigi 329-04		
	Japan		
	Phone: 0285-44-2111 x3334		
	Fax: 0285-44-1557.		
FEATURES	Location/Qualifiers		
source	1..1863		
	/organism="Hepatitis C virus"		
	/isolate="HC-J2"		
	325..>1863		
	/codon_start=1		
	/product="structural protein"		
	/db_xref="PID:d1001434"		
	/db_xref="PID:g221589"		
	/translation="MSTNPKQKTKNTNRRPDVKFGCGQIVGCVYLPGRGRPL GVRATKTSQSPGRORIPKARPEGRWAQPGYPWPYLGNEGLGWAAGLLSPRG SRPSWGPTPPRRSRNLKGVITDITLCFADLMGYIPLVGAPLGGAARALAHGVRVLE SYNVAQNLPGCSSTIFLALLSLCTIPASAYEVNRVNSGVYHVTNDCSNSSIVYEAAQ LIMVATGCGVTVCRNNSSRCVWALPTPLAANTSIPTTTIRRHVDLLVGAAATCSAM		
CDS			

VGDLCGSFVLSQFTSPRRHETLQDNCISYPGHLSGHRMAWDMWMMNSPTTALVW
SOLLRIPTOTVDMVTGAHGVLAGLAYYSVGNWAKVLIYVLLFAGVDGTHVTGAT
GHTFSGIASLFLPGASOKIOLINTNGSHNRTALNCNDSINTGFLAALFYTHKFNAS
GCPERLASCRSDIGFDQGMGPITYTEPQDSQDKPYCHMYAPQRCSWSAADVCGPVYC
FTPSP*

BASE COUNT 344 a 574 c 539 g 406 t
ORIGIN

Query Match 89.3%; Score 283; DB 122; Length 1863;
Best Local Similarity 90.5%; Pred. No. 4.36e-214;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Caps 0;

Db 325 atgagcagcaatcctaaacctcaagaagaaacaaacgtacacacacacgtcgccacag 384
|||||
|||||

Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACAGCTAACACACCGCGCCCTCAK 60
|||||

Db 385 gacgtcaagttcccgggcgcggtcgatcggtggagtttacctgttgcgcgcag 444
|:|||
|||||

Qy 61 GGSCTNNNNNNCCGGTGGCGGTGAGATCGTGTGGAGTTTACCTGTTGGCGCAGG 120
|||||

Db 445 gqccccaggttggtggtgcgcgcgactaggaagacttcgagcgtgcgaacctcgtgga 504
|||||
|||||

Qy 121 GGCCCCAGGNNGGTGTGGCGCGACTAGGAAGACTTCGCGCGGTCAACCTCGTGGC 180
|||||
|||||

Db 505 agggcgaacctatccccaggtccggtccgagcgcagcgtggcctgagccgcgcgg 564
|||||
|||||

Qy 181 AGCGCAGACGCTATCCCAAGGCTGCGGCGCGAGCGAGCTCGGCTGACGCCGGG 240
|||||
|||||

Db 565 taacctggccctctatgcaatagggtcagggtggcaggaagctcctgctcaccc 624
||
|||||

Qy 241 TATCTTGGCCCTCTATGGAATGAGGCTCGGCTGGCGCGGTGCTGCTGCTCCGCC 300
|||||
|||||

Db 625 cgtggctcgtcgctagttggggcccc 651
||
|||||

Qy 301 CGCGGCTCGCGCCCAATGGGGCCCC 327
||
|||||

RESULT 14 HPCK1R2 9410 bp ss-RNA VRL 02-FEB-1997
LOCUS Hepatitis C virus (strain HCV-1b, clone HCV-K1-R2), complete genome
DEFINITION sequence.
ACCESSION D50481
NID g1030705
KEYWORDS polyprotein; core protein; E1; E2; NS2; NS3; NS4A; NS4B; NS5A;
NS5b; envelope protein; non-structural protein;
interferon-sensitive; interferon-resistant; IFN-sensitive;
IFN-resistant; ISDR; interferon sensitivity determining region;
HVR; hypervariable region.
SOURCE Hepatitis C virus (strain:HCV-1b) cDNA to genomic RNA,
clone:HCV-K1-R2.
ORGANISM Hepatitis C virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Flaviviridae;
Hepatitis C-like viruses.
1 (sites)

REFERENCE
AUTHORS Enomoto,N., Sakuma,I., Asahina,Y., Kurosaki,M., Murakami,T.,
Yamamoto,C., Izumi,N., Marumo,F. and Sato,C.
TITLE Comparison of full-length sequences of interferon-sensitive and
resistant hepatitis C virus 1b. Sensitivity to interferon is
conferred by amino acid substitutions in the NS5A region
J. Clin. Invest. 96 (1), 224-230 (1995)
95340824

REFERENCE
MEDLINE 2 (bases 1 to 9410)
AUTHORS Enomoto,N.
JOURNAL Unpublished (1995)

REFERENCE 3 (bases 1 to 9410)

AUTHORS Enomoto,N.

DIRECT SUBMISSION

TITLE Submitted (08-MAY-1995) to the DDBJ/EMBL/GenBank databases.

JOURNAL Nobuyuki Enomoto, Tokyo Medical and Dental University, Second
Department of Internal Medicine; 1-5-45 Yushima, Bunkyo-ku, Tokyo
113, Japan (E-mail:PXN04522@niftyserve.or.jp,
Tel:03-3813-6111(ex.3224), Fax:03-3818-7177)

FEATURES Location/Qualifiers

1..9410

source

/organism="Hepatitis C virus"

/note="IFN-resistant HCV from patient 2"

/strain="HCV-1b"

/clones="HCV-K1-R2"

330..9362

/codon_start=1

/product="polyprotein"

/db_xref="PID:d1009707"

/db_xref="PID:g1814086"

/translation="MSINPQRKTKNTNRPRQDVKEPGGGQIVGGYLLPRGRRL
GVRAIRKTSERSQRRQIPKARQEGRAMAQGYWPLYNGEGMGAGWLLSPRG
SRPSKGTDPRRRSNRLVIDTLCGFADLMCYIPIVGAFLGASRALAHGVRLD
GVNATGNLPGCSFIFLALLSLCLTPASAEVVRNASGVYHVTNDCNSASIVYEAD
MIHTPGCVPCVREANSRCWALPTTLAARNSSVPTTIRRHVDLLVGAATCSAMY
VDLCSGFVLSQFTSPRRHETVQDNCISYPGHVSGRMAWDMWMMNSPTTALVW
SOLLRIPTQAVMDVAGAHGVLAGLAYYSVGNWAKVLIYVLLFAGVDGNTYVSGGNK
SGPERIASCRSIDAFAGQMGPIYAEPRGSSDQRPYCHMYAPRPGIIPASEVCGPVYC
SHITGGLSIFAPGAQKQIQLVNTGSHNRTALNCNDSINTGFLAALFYTHFNSS
FTSPVVVGTDRSCAPTYSWGENETDVLINTRAQGNWECCTMNGTGTCTKCGG
PPCNIIGVGNNTLCPTCFRKHPEATYTKCGSPWLTPRCWVDYPIYRLHYPCTVNF
TIFKRVYVGGVHRLNAQNRCTGREDRSELSPLLSITENQILPISFTTL
PVLSTGLIHLHONIVDQVLYGVSSAVSVFIKWEYIILLLFLLADARVACALWMLL
IAQAAALENLVVLAASVAGAHGFLSIVFFCAAWYIKGLVPGAAAYFYGWPLLL
LILLALPRAYAMDREMAACGGAGVEFGLALLTSLPHYKVYIARLIWLIQYITKAPYVR
LQWIPPLNVGRGRDAIILITCALHPELIFSITKILVALIPLVLAOGITKAPYFVR
AQGLIRACHLVKRVAGHYVQMAFKLAALTGTYYVYDHLTPLDQNAHTGLDLAVNE
PVVSDMETKIIITWADTACGDIILGLAVSARRKEIFLGPADSLGSGWRLLAPIT
AIPIETIKGRRHIFCHSKKCGDELAARLGLGINAVAYYRGIDVSVIPASGDVVVA
TIALMTGTCDFSDVIDCNTVTQVDFSLDPTETIETTVODAVSRORRGTGRG
RAGIYRVPGERSGFDSVLCEDDACCAMYLEIPAEYSVRLRAYLNTPLGVPWQ
DLHEFVESYVTLGTHDAHFLSOTKOAGDNFYLVAOATVCAARAQAPPSWDMKKC
LIRLKPTLHGPTLLYRLGAVQNEVILTHPITKYIMACSADLEVVYTSWLVGGVLA
ALAAVCLTGSVWVIGRIILSGKPAIIPREVLYREFDEMEECASHLPYIEQCQIAE
QFKQALIGLIQTATQAEAAAPVWESQALETFWAKHMNFISGIQYLAGLSITLPGN
PAIASLMAFTASITPLITTOHTLFFNILGQWAAOLAPPAASAFVAGAGAGAGS1
GLCKVVDVILAGYAGVAGALVAFKVMGSDMPSTEDLVNLLPALLSPCALVVGWCAA
ILRRHVCGEGAVQMMRLIFASGHNVSPHIVVPESDAARVQTILSSLTITLQIK
PLHKWINEDCSTPCSGSLRDVDMICITVLSDFKTMQLSRVPLRPVLPVPSLQCRQYK
GYWPGDGIWQTTCPGGAQITGHVKNCSMRIVGPKTCNTWHTGTFINAYTTGCTPSP
APNYSRLARVAAEEYVEVTRGVYVGTMTDNKCPQVPAPEFFKELDQVRLRR
YAPASRPLRDEVTFQVGLNVQVGSQLPCEPEPDVALVLSMTDPSHITAEAKRLR
ARGSPSSASSQSLSKATCTTHDSDADLLEANLLWROEMGNIITVSEEN
KVVILDSFPLAEDEREVSLPARIILKSKRFPAMPINWAPDINPPLIESKHQDPT
VPPVHCCPLPTKAPPIPPRRRRTVYTESTVSSALAEATKTCGSSSSAAQSDY
ATAPPDQSSDGDGDSVESYSNPELEGEPDPLSDGSGSWTSVEEAGDWCCSMS
YTWGALITPAAESKLPINALSNSILRHNNVYATTSRSASQRQKVTEDRLVD
DHYRDVLEKAKASTVKARLLSVEEAKLTPPHSARSKFYGKADVRNLSRAINHI
RSWKDLQEDTETPTDITIMAKNEVFCVQPEKGGKPARLIVFPDLGVRCXKMAIYD

Result No.	Score	Query		DB ID	Description	Pred. No.	
		Match	Length				
1	317	100.0	327	25	T27937	Hepatitis C virus	5,34e-189
2	285	89.9	573	21	T16620	Hepatitis C virus	1.16e-167
3	285	89.9	573	21	T16621	Hepatitis C virus	1.16e-167
4	283	89.3	327	25	T27939	Hepatitis C virus	2.49e-166
5	283	89.3	570	3	Q21747	Hepatitis C virus	2.49e-166
6	283	89.3	573	21	T16617	Hepatitis C virus	2.49e-166

CC The sequences T27937-T27989 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/E1, NS4 or NS5B regions of the genome.
CC This sequence represents nucleotides 1-310 from the HCV type 1d isolate
CC BNL1.

CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/E1 and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to synthesise probes and primers for the
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 327 BP; 56 A; 102 C; 104 G; 51 T;

Query Match 100.0%; Score 317; DB 25; Length 327;

Best Local Similarity 100.0%; Pred. No. 5,34e-189;

Matches 327; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Db 1 atgagcacgaatcctaaacctcaaaagaaacacaaacgtacacacacccgcgcctctak 60
|||||
Qy 1 ATGAGCAGCAATCCTTAACCTCAAAAGAAACCAACGCTACACCAACCGCGCCTCAK 60
|||||

Db 61 gqgctnnnnnnccgggtggcggtcagatcggtggagttacctgttgcgcgcagg 120
|||||
Qy 61 GSGTNNNNNNCCGGGTGGCGGTCAAGTCGTTGGTGAGTTTACCTGTTGCCGCGAGG 120
|||||

Db 121 ggcacacagmggtgtgcgcgcactaggaagacttcgagcggttcacacactcgtggc 180
|||||
Qy 121 GCGCCAGGNNGGGTGCGCGCGACTAGGAAGACTTCGAGCGGTCACAACTCGTGGC 180
|||||

Db 181 aggcagacactatccccaggtctcgcgcgcaggtcgcgcgcaggtcctggctcagccggg 240
|||||
Qy 181 AGGCAGACGCTATCCCCAGGCTCGCGCGAGGTCGCGCGAGGCTCCTGGGCTCAGCCGGG 240
|||||

Db 241 taccctggccctctatggcaatgaggtgaggtggcggtggcggtggcggtgctgccccc 300
|||||
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTGCGGCTGCGGCTGGGNTGGTCTGTCCCC 300
|||||

Db 301 cgtggtctcggcccaattggggcccc 327
|||||
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327
|||||
```

RESULT 2

ID T16620 standard; cDNA; 573 BP.
AC T16620;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate P10 core protein gene.
KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 1..573
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.

PR 15-AUG-1994; US-290665.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI; 96-139709/14.
DR P-PSDB; R92946.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
PS Claim 3; Page 156; 340pp; English.
CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
CC isolates. The isolated sequences are useful for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
CC used in the prevention of HCV infection.
SQ Sequence 573 BP; 96 A; 185 C; 179 G; 113 T;

Query Match 89.9%; Score 285; DB 21; Length 573;

Best Local Similarity 90.8%; Pred. No. 1.16e-167;

Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

```
Db 1 atgagcacgaatcctaaacctcaaaagaaacacaaacgtacacacacccgcgcacag 60
|||||
Qy 1 ATGAGCAGCAATCCTTAACCTCAAAAGAAACCAACGCTACACCAACCGCGCCTCAK 60
|||||

Db 61 gacgtcaagttcccgcggtggcggtcagatcggtggagttacctgttgcgcgcagg 120
|||||
Qy 61 GSGTNNNNNNCCGGGTGGCGGTCAAGTCGTTGGTGAGTTTACCTGTTGCCGCGAGG 120
|||||

Db 121 ggcacacagmggtgtgcgcgcactaggaagacttcgagcggttcacacactcgtggc 180
|||||
Qy 121 GCGCCAGGNNGGGTGCGCGCGACTAGGAAGACTTCGAGCGGTCACAACTCGTGGC 180
|||||

Db 181 aggcagacactatccccaggtctcgcgcgcaggtcgcgcgcaggtcctggctcagccggg 240
|||||
Qy 181 AGGCAGACGCTATCCCCAGGCTCGCGCGAGGTCGCGCGAGGCTCCTGGGCTCAGCCGGG 240
|||||

Db 241 taccctggccctctatggcaatgaggttgggtggcggtggcggtggcggtgctgccccc 300
|||||
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTGCGGCTGCGGCTGGGNTGGTCTGTCCCC 300
|||||

Db 301 cgtggtctcggcccaattggggcccc 327
|||||
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327
|||||
```

RESULT 3

ID T16621 standard; cDNA; 573 BP.
AC T16621;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate DK1 core protein gene.
KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 1..573
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.

PA (CHIR-) CHIRON CORP.
PI Houghton M, Choo QL, Kuo G, Weiner AJ, Urdea MS, Irvine BD;
PI Kolberg JA;
DR WPI; 92-080094/10.
PT Reagents for isolating, amplifying and detecting HCV
PT polynucleotide(s) - used to monitor spread of blood-borne non-a,
PT non-b hepatitis virus infection and screen blood samples for
PT virus
PS Disclosure; Fig 2; 6pp; English.
CC The sequences represented in Q21746-50 are DNA consensus
CC sequences for five different HCV isolates from different
CC geographic locations (Japan and U.S.).
SQ Sequence 570 BP; 94 A; 183 C; 177 G; 116 T;

Query Match 89.3%; Score 283; DB 3; Length 570;
Best Local Similarity 90.5%; Pred. No. 2.49e-166;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db 1 atgagcacgaatcctaaacctcaaaagaaacaaacgaacacacacacgctgcgccacag 60
Qy 1 ATGAGCACGAATCCTAAACCTCAAGAAACCAACACGTAACACCAACCGCGCCCTCAK 60
Db 61 gagctcaagttcccggtggtgcgtcagatcgttggtagttacttctgtgcgcgcagg 120
Qy 61 GCGGTNNNNNNNCGGCGTGGCGGTGAGATCGTGTGGAGTTTACCTGTTCGCCGCGAGG 120
Db 121 ggcctcaggttgggtgtgcgcgcaggaagacttccgagcgttcgcaacctcgaggt 180
Qy 121 GGCCCCAGGNNGGGTGTGCGCGGCACTAGGAAGACTTCGAGCGGTCAACACTCGTGGC 180
Db 181 agactcagcctatccccaggaagcgtgcgcgcagggcaggaactgggtcagccggg 240
Qy 181 AGCGGACAGCCTATCCCGAGGCTGCGYCGGCGGAGGCGAGGCTCTGGGCTACGCCGGG 240
Db 241 tatcttgcccctctatggcaatgaggtcgggtggcggtggtcctctctctccc 300
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGCGGCGGAGTGGCTCTGTCCCCC 300
Db 301 cgtggtctcgccctagctggggcccc 327
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 6
ID T16617 standard; cDNA; 573 BP.
AC T16617;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate S45 core protein gene.
KW HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 1..573
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
PN W09605315-A2.
PD 22-FEB-1996.
PE 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI; 96-139709/14.

DR P-PSDB; R92943.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
PS Claim 3; Page 154; 340pp; English.
CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
CC isolates. The isolated sequences are useful for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs
CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
CC used in the prevention of HCV infection.
SQ Sequence 573 BP; 95 A; 189 C; 177 G; 112 T;

Query Match 89.3%; Score 283; DB 21; Length 573;
Best Local Similarity 90.5%; Pred. No. 2.49e-166;
Matches 296; Conservative 4; Mismatches 27; Indels 0; Gaps 0;
Db 1 atgagcacgaatcctaaacctcaaaagaaacaaacgaacacacacacgctgcgccacag 60
Qy 1 ATGAGCACGAATCCTAAACCTCAAGAAACCAACACGTAACACCAACCGCGCCCTCAK 60
Db 61 gagctcaagttcccggtggtgcgtcagatcgttggtagttacttctgtgcgcgcagg 120
Qy 61 GCGGTNNNNNNNCGGCGTGGCGGTGAGATCGTGTGGAGTTTACCTGTTCGCCGCGAGG 120
Db 121 ggcctcaggttgggtgtgcgcgcagtaggaacttccgagcgttcacaaacctcgtga 180
Qy 121 GGCCCCAGGNNGGGTGTGCGCGGCACTAGGAAGACTTCGAGCGGTCAACACTCGTGGC 180
Db 181 cggcagcaactatccccaggtcgcgcgcagggcaggaagcgtgggcccagccggg 240
Qy 181 AGCGGACAGCCTATCCCGAGGCTGCGYCGGCGGAGGCGAGGCTCTGGGCTACGCCGGG 240
Db 241 catcttgcccctctatggcaatgaggtcgttgggtggcgaggtggtcctctctcacc 300
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGCGGCGGAGTGGCTCTGTCCCCC 300
Db 301 cgtggtctcgccctagctggggcccc 327
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 7
ID Q38959 standard; cDNA; 9391 BP.
AC Q38959;
DT 29-JUL-1993 (first entry)
DE Hepatitis C virus gene.
KW HCV; detection; antigen; vaccine; recombinant; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 305..9349
FT /*tag= a
FT /product= a
FT /note= "does not contain stop codon"
PN J05068563-A.
PD 23-MAR-1993.
PE 17-JUL-1991; 203884.
PR 17-JUL-1991; JP-203884.
PA (KAGA) KAGAKU OYOBI KESSEI RYOHQ.
DR WPI; 93-130639/16.
DR P-PSDB; R35207.
PT Nucleotide sequence encoding hepatitis C virus polypeptide - is
PT useful for detecting HCV infection, esp. in Japan, and as vaccine
PT against HCV
PS Claim 1; Page 6-17; 17pp; Japanese.

CC development. Antibodies directed against it can be used for
CC screening antiviral agents and for isolation of non-A, non-B hepat-
CC itis (NANBH).
CC See also Q11076-79.
SQ Sequence 552 BP; 98 A; 161 C; 176 G; 103 T;

Query Match 88.6%; Score 281; DB 2; Length 552;
Best Local Similarity 88.1%; Pred. No. 5,34e-165;
Matches 288; Conservative 11; Mismatches 28; Indels 0; Gaps 0;

Db 91 atgagcacaatcctaaaccysaaagaaacaaacaaacgtacacacacacgtcgccacag 150
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :||||| :
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACACGTACACCAACCGCGCCTCAK 60

Db 151 gacgtvaagtccckgpcggtggtcgagatcgtygggtgagttactgttgccgcaggg 210
|:| ||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 61 GSGTNNNNNNCCGGTGGCGGTGAGATCGTGTGGAGTTTACCTGTTGCCGCGCAGG 120

Db 211 gggccaggttggtggtggtgagactagaaagacttcgagcggtcraacactcgtgga 270
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 121 GGGCCCGAGGNNGGGTGTCGGCGGCACTAGCAAGACTTCCGAGCGGTCAACAACCTCGTGGC 180

Db 271 aggcgcacacatccccaaagctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 330
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 181 AGCGACAGCCTATCCCAAGGCTCGYCGYCGGCGGCGGCGGCGGCTCGGCTCAGCGCGG 240

Db 331 tatecttgccctctatggcaatgaggctwgggtgggcaggatggctctctcgcaccc 390
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGCGGNTGGCTCTGTGCCCC 300

Db 391 cgcgctctcgccctagttggggcccy 417
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 301 CGCGGCTCTCGGCGCAATTGGGGCCCC 327

RESULT 10

ID T16618 standard; cDNA; 573 BP.
AC T16618;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate D1 core protein gene.
KW HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 1..573
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI; 96-139709/14.
DR P-PSDB; R92944.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
PS Claim 3; Page 155; 340pp; English.
CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
CC isolates. The isolated sequences are useful for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The
CC proteins may also be used to detect antibodies against HCV in serum,
CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
CC used in the prevention of HCV infection.
SQ Sequence 573 BP; 96 A; 184 C; 177 G; 116 T;

Query Match 88.6%; Score 281; DB 21; Length 573;
Best Local Similarity 90.2%; Pred. No. 5,34e-165;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcacaatcctaaacag 60
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAACCAACACGTACACCAACCGCGCCTCAK 60

Db 61 gacgtvaagtccckgpcggtggtcgagatcgtygggtgagttactgttgccgcaggg 120
|:| ||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 61 GSGTNNNNNNCCGGTGGCGGTGAGATCGTGTGGAGTTTACCTGTTGCCGCGCAGG 120

Db 121 gggccaggttggtggtggtgagactagaaagacttcgagcggtcraacactcgtgga 180
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 121 GGGCCCGAGGNNGGGTGTCGGCGGCACTAGCAAGACTTCCGAGCGGTCAACAACCTCGTGGC 180

Db 181 aggcgcacacatccccaaagctccgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgcgc 240
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 181 AGCGACAGCCTATCCCAAGGCTCGYCGYCGGCGGCGGCGGCTCGGCTCAGCGCGG 240

Db 241 tatecttgccctctatggcaatgaggctwgggtgggcaggatggctctctcgcaccc 300
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTCGCGGTGGCGGNTGGCTCTGTGCCCC 300

Db 301 cgcgctctcgccctagttggggcccy 327
||||| ||||||| :||||| :||||| :||||| :||||| :||||| :
Qy 301 CGCGGCTCTCGGCGCAATTGGGGCCCC 327

RESULT 11

ID T16622 standard; cDNA; 573 BP.
AC T16622;
DT 01-OCT-1996 (first entry)
DE Hepatitis C virus isolate T10 core protein gene.
KW HCV; El; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis; ss.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT CDS 1..573
FT /*tag= a
FT /product= core_protein
FT /note= "does not contain stop codon"
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH) US SEC DEPT HEALTH.
PI Bukh J, Miller RH, Purcell RH;
DR WPI; 96-139709/14.
DR P-PSDB; R92944.
PT DNA and amino acid sequence of HCV envelope 1 and core proteins -
PT used to determine HCV genotype and as vaccines against HCV infection
PS Claim 3; Page 157; 340pp; English.
CC T16610-T16661 are cDNAs encoding a core protein gene of 52 HCV
CC isolates. The isolated sequences are useful for the prodn. of primers
CC useful for detecting the presence of HCV in a sample, the primers
CC are also useful for HCV genotyping. Proteins encoded by the cDNAs

CC can be used in vaccines for immunising against HCV infection. The
 CC proteins may also be used to detect antibodies against HCV in serum,
 CC saliva, lymphocytes or other mononuclear cells. The antibodies may be
 CC used in the prevention of HCV infection.

SQ	Sequence	573 BP;	97 A;	185 G;	177 G;	114 T;
----	----------	---------	-------	--------	--------	--------

Query Match	88.6%	Score 281;	DB 21;	Length 573;
Best Local Similarity	90.2%;	Pred. No. 5.34e-165;		
Matches	295;	Conservative	4;	Mismatches 28; Indels 0; Gaps 0;

Db 1 atgagcacgaatcctaaactcaagaaaaaaccaacgtaacacaaacgcgcgcacag 60
|||||
Qy 1 ATGACACGAATCCTTAACCTCAAAGAAAAACCAACCGTAAACACCAACCGCGGCCTCAK 60

Db 61 gacgtcaagtcccggcgggtggccagatcgttggtgagtttacctgttcgcgcgcagg 120
| : | ||||| || ||||||| ||||||| ||||||| ||||||| |||||||
Ov 61 GGSGTNNNNNCCGGGTGGCGGTCAGATCGTTCGTGGAGCTTTACCTGTTGCCCGCAGG 120

Qy 61 GSGTNNNNNNCCGGGTGGCGGTGAGTCGTTGGTGGAGCTTTACCTGTTGCCGCCGAGG 120

Db 121 ggccccagggttggtgtgcgcgcgactaggaagacttcgagcgggtcgcaactcgtgga 180

Db 121 ggcgccaggtgggtgctgcgcgcactaggaagactccgagcggctcgcaactcgtgga 180
|||||
Qy 121 ggcgccagnnngggtgctgcgcgcactaggaagactccgagcgggtcacaaacctcgtggc 180

181 AGCGACACCTATCCCAAGGCTCGCGGCGGAGGCTCTGGGCTAGCCCCGG 240

[illegible]

Db	301	cgggctcgcgctaagtgggcccc	327
Qy	301	CGGGCTCTCGGCCAATTGGGCCC	327

RESULT 15
ID Q44921 standard; DNA; 686 BP.
AC Q44921;

DT 04-OCT-1994 (first entry)
DE Hepatitis C virus 5'-untranslated region and part of core region.
KW Hepatitis C virus; HCV; non-A, non-B hepatitis virus; NANBHV;

KW antisense oligonucleotide; translation inhibition; therapy;
KW 5'-untranslated region; 5'-UTR; hybridisation target; ss.
OS Hepatitis C virus.

EH	Key	location/Qualifiers
FT	5'UTR	1..341
FT	/*tag= a	
FT	cons	302..686

```

FT      CDS      347..686
FT      /*tag= b
FT      /product= Core_protein
FT      /note= "partial coding region".

```

PN W09405813-A.
PD 17-MAR-1994.
PF 10-SEP-1993; J01293.

PR 10-SEP-1992; US-945289.
PR 14-APR-1993; JP-087195.
PA (KAGA) CHEMO SERO THERAPEUTIC RES INST.

PA (ISIS-) ISIS PHARM INC.
PA (MOCH) MOCHIDA PHARM CO LTD.
PI Anderson KP, Eto T, Furukawa S, Hamada F, Hanecak RC;

PI Hoshito K, Nakatake H, Nishihara T, Nozaki C;
DR WPI; 94-101217/12.
PT Anti:sense oligo:nucleotide(s) complementary to hepatitis C viral

PT genome - useful for inhibiting HCV replication, to treat related

RESULT 15
ID Q44921 standard; DNA; 686 BP.
AC Q44921;

DT 04-OCT-1994 (first entry)
DE Hepatitis C virus 5'-untranslated region and part of core region.
KW Hepatitis C virus; HCV; non-A, non-B hepatitis virus; NANBHV;

KW antisense oligonucleotide; translation inhibition; therapy;
KW 5'-untranslated region; 5'-UTR; hybridisation target; ss.
OS Hepatitis C virus.

EH	Key	location/Qualifiers
FT	5'UTR	1..341
FT	/*tag= a	
FT	cns	302..686

```

FT      CDS      347..686
FT      /*tag= b
FT      /product= Core_protein
FT      /note= "partial coding region".

```

PN W09405813-A.
PD 17-MAR-1994.
PF 10-SEP-1993; J01293.

PR 10-SEP-1992; US-945289.
PR 14-APR-1993; JP-087195.
PA (KAGA) CHEMO SERO THERAPEUTIC RES INST.

PA (ISIS-) ISIS PHARM INC.
PA (MOCH) MOCHIDA PHARM CO LTD.
PI Anderson KP, Eto T, Furukawa S, Hamada F, Hanecak RC;

PI Hoshito K, Nakatake H, Nishihara T, Nozaki C;
DR WPI; 94-101217/12.
PT Anti:sense oligo:nucleotide(s) complementary to hepatitis C viral

PT genome - useful for inhibiting HCV replication, to treat related

Job time : 64 secs.

工
L
回
L
回
Z

(MT)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch nn n.a. - n.a. database search, using Smith-Waterman algorithm

Run on: Mon Feb 23 21:01:42 1998; MasPar time 23.30 Seconds
645.468 Million cell updates/sec

Tabular output not generated.

Title: >US-08-836-075A-1

Description: (1-327) from US08836075A.seq

Perfect Score: 317

N.A. Sequence: 1 ATGAGCAGCAATCCTAAACC.....CTCGGCCCAATTGGGGCCCC

Scoring table: TABLE default

Gap 6

Nmatch STD : Dbase 0: Query 0

Searched: 87531 seqs, 22996021 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: n-issued

1:back1 2:51 3:52 4:53 5:54 6:55 7:56 8:PT90 9:PT91
10:PT92 11:PT93 12:PT94 13:PT95 14:PT96

Statistics: Mean 7.562; Variance 4.432; scale 1.706

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description	Pred. No.
		Match	Length				
1	285	89.9	573	13	PCT-US95-1	Sequence 114,	Applicat 2.94e-183
2	285	89.9	573	13	PCT-US95-1	Sequence 113,	Applicat 2.94e-183
3	283	89.3	573	13	PCT-US95-1	Sequence 110,	Applicat 8.21e-182
4	281	88.6	573	13	PCT-US95-1	Sequence 117,	Applicat 2.29e-180
5	281	88.6	573	13	PCT-US95-1	Sequence 118,	Applicat 2.29e-180
6	281	88.6	573	13	PCT-US95-1	Sequence 115,	Applicat 2.29e-180
7	281	88.6	573	13	PCT-US95-1	Sequence 111,	Applicat 2.29e-180
8	279	88.0	573	13	PCT-US95-1	Sequence 112,	Applicat 6.39e-179
9	279	88.0	573	13	PCT-US95-1	Sequence 105,	Applicat 6.39e-179
10	279	88.0	573	13	PCT-US95-1	Sequence 124,	Applicat 6.39e-179

11	279	88.0	573	13	PCT-US95-1	Sequence 120	Applicat	6.39e-179
12	279	88.0	573	13	PCT-US95-1	Sequence 108	Applicat	6.39e-179
13	279	88.0	573	13	PCT-US95-1	Sequence 121	Applicat	6.39e-179
14	278	87.7	573	13	PCT-US95-1	Sequence 106	Applicat	3.37e-178
15	277	87.4	573	13	PCT-US95-1	Sequence 123	Applicat	1.78e-177
16	277	87.4	573	13	PCT-US95-1	Sequence 104	Applicat	1.78e-177
17	277	87.4	669	7	US-08-143	Sequence 3	Applicatio	1.78e-177
18	277	87.4	669	7	US-08-143	Sequence 3	Applicatio	1.78e-177
19	277	87.4	9401	9	PCT-US91-0	Sequence 9	Applicatio	1.78e-177
20	275	86.8	573	13	PCT-US95-1	Sequence 116	Applicat	4.95e-176
21	275	86.8	573	13	PCT-US95-1	Sequence 122	Applicat	4.95e-176
22	275	86.8	573	13	PCT-US95-1	Sequence 103	Applicat	4.95e-176
23	275	86.8	573	13	PCT-US95-1	Sequence 119	Applicat	4.95e-176
24	275	86.8	573	13	PCT-US95-1	Sequence 109	Applicat	4.95e-176
25	274	86.4	573	13	PCT-US95-1	Sequence 107	Applicat	2.61e-175
26	273	86.1	462	13	PCT-US95-1	Sequence 1	Applicatio	1.38e-174
27	273	86.1	9401	7	US-07-910	Sequence 9	Applicatio	1.38e-174
28	271	85.5	657	9	PCT-US91-0	Sequence 11	Applicati	3.83e-173
29	271	85.5	657	5	US-07-681	Sequence 11	Applicati	3.83e-173
30	271	85.5	923	13	PCT-US95-1	Sequence 14	Applicati	3.83e-173
31	270	85.2	3075	7	US-07-910	Sequence 11	Applicati	2.00e-172
32	267	84.2	573	13	PCT-US95-1	Sequence 142	Applicat	2.96e-169
33	265	83.6	360	9	PCT-US91-0	Sequence 17	Applicati	8.20e-169
34	265	83.6	360	5	US-07-681	Sequence 17	Applicati	8.20e-169
35	265	83.6	453	5	US-07-681	Sequence 15	Applicati	8.20e-169
36	265	83.6	453	9	PCT-US91-0	Sequence 15	Applicati	8.20e-169
37	265	83.6	657	9	PCT-US91-0	Sequence 13	Applicati	8.20e-169
38	264	83.3	657	5	US-07-681	Sequence 13	Applicati	8.20e-169
39	264	83.3	657	13	PCT-US95-1	Sequence 139	Applicat	4.32e-168
40	264	83.3	573	13	PCT-US95-1	Sequence 145	Applicat	4.32e-168
41	260	82.0	573	13	PCT-US95-1	Sequence 140	Applicat	3.32e-165
42	257	81.1	573	13	PCT-US95-1	Sequence 141	Applicat	4.83e-163
43	256	80.8	573	13	PCT-US95-1	Sequence 143	Applicat	2.54e-162
44	247	77.9	573	13	PCT-US95-1	Sequence 151	Applicat	2.14e-154
45	245	77.3	573	13	PCT-US95-1	Sequence 150	Applicat	2.14e-154

ALIGNMENTS

RESULT	1	
ID	PCT-US95-10398-114	STANDARD; DNA; UNC; 573 BP.
AC	xxxxxx	
DT	01-JAN-1900	
DE	Sequence 114,	Application PC/TUS9510398.
CC	Sequence 114,	Application PC/TUS9510398
CC	GENERAL INFORMATION:	
CC	APPLICANT:	BUKH, J., MILLER, R.H. AND
CC	APPLICANT:	PURCELL, R.H.
CC	TITLE OF INVENTION:	NUCLEOTIDE AND DEDUCED
CC	TITLE OF INVENTION:	AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC	TITLE OF INVENTION:	CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC	TITLE OF INVENTION:	AND THE USE OF REAGENTS DERIVED FROM THESE
CC	TITLE OF INVENTION:	SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC	NUMBER OF SEQUENCES:	263
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE:	MORGAN & FINNEGAN
CC	STREET:	345 PARK AVENUE
CC	CITY:	NEW YORK
CC	STATE:	NEW YORK
CC	COUNTRY:	USA
CC	ZIP:	10154
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE:	FLOPPY DISK
CC	COMPUTER:	IBM PC COMPATIBLE

CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 114:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: DK1
SQ Sequence 573 BP; 98 A; 185 C; 177 G; 113 T; 0 other;

Query Match 89.9%; Score 285; DB 13; Length 573;
Best Local Similarity 90.8%; Pred. No. 2.94e-183;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Dbb 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
Qy 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTCAK 60
|||||
Dbb 61 GACGTCAAGTTCCGGCGGTGTCAGATGTTGTCGAGTTTACCTGTGCGCGCAGG 120
|:|:|
Qy 61 GCGGTNNNNNNCCGGTGGCGGTGTCAGATGTTGTCGAGTTTACCTGTGCGCGCAGG 120
|:|:|
Dbb 121 GCGCCAGGTTGGTGTGCGCGGCTAGCAAGACTTCCGAGCGGTGCGCAACCTCTGTGA 180
|||||
Qy 121 GCGCCAGGTTGGTGTGCGCGGCTAGCAAGACTTCCGAGCGGTGCGCAACCTCTGTGC 180
|||||
Dbb 181 AGCGACACCTATCCCAAGGCTCGCGGCGCGGCGGCGGCTGCGGCTCAGCCCGGG 240
|||||
Qy 181 AGCGACACCTATCCCAAGGCTCGCGGCGCGGCGGCGGCTGCGGCTCAGCCCGGG 240
|||||
Dbb 241 TACCTTGGCCCTCTATGGCATGAGGCAATGGGCTGGCGGAGATGGCTCTGTCAACC 300
|||||
Qy 241 TATCTTGGCCCTCTATGGCATGAGGCTGGGCGGAGATGGCTCTGTCTGTCCGCC 300
|||||
Dbb 301 CGGGCTCTGGGCTAGTTGGGCGCC 327
|||||
Qy 301 CGGGCTCTGGGCGCAATGGGCGCC 327
|||||

RESULT 2
ID PCT-US95-10398-113 STANDARD; DNA; UNC; 573 BP.
AC xxxxxx
DT 01-JAN-1900

DE Sequence 113, Application PCT/US95/10398.
CC Sequence 113, Application PCT/US95/10398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 113:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: P10
SQ Sequence 573 BP; 96 A; 185 C; 179 G; 113 T; 0 other;

Query Match 89.9%; Score 285; DB 13; Length 573;
Best Local Similarity 90.8%; Pred. No. 2.94e-183;
Matches 297; Conservative 4; Mismatches 26; Indels 0; Gaps 0;

Dbb 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCACAG 60
|||||
Qy 1 ATGACGACGAATCTTAACCTCAAGAAACCAACGTAACACCAACCGCCGCCCTCAK 60
|||||
Dbb 61 GACGTCAAGTTCCGGCGGTGTCAGATGTTGTCGAGTTTACCTGTGCGCGCAGG 120
|:|:|
Qy 61 GCGGTNNNNNNCCGGTGGCGGTGTCAGATGTTGTCGAGTTTACCTGTGCGCGCAGG 120
|:|:|


```
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 117:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND3
CC Sequence 573 BP; 100 A; 184 C; 174 G; 115 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db      1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy      1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db      61 GAGCTCAAGTTCCGGCGGTGCGGAGATCGTTGGTGAGTTTACCTGTCGCCGCCAGG 120
Qy      61 GSGGTNNNNNNCCGGTGGCGGTGAGATCGTTGGTGAGTTTACCTGTCGCCGCCAGG 120

Db      121 GCGCCAGGTTGGTGTGCGCGGCGACTAGGAAGACTTCCGAGCGGTGCGCAACCTCGTGA 180
Qy      121 GCGCCAGGTTGGTGTGCGCGGCGACTAGGAAGACTTCCGAGCGGTGCGCAACCTCGTGGC 180

Db      181 AGCGCAGCAACTATCCCAAGGCTCGCGCGCGGAGGTAGGGCTGAGGGCTCAGCCCGGG 240
Qy      181 AGCGCAGGCTATCCCAAGGCTCGCGCGCGGAGGTAGGGCTGAGGGCTCAGCCCGGG 240

Db      241 TACCGCTTGGCCCTCTATGGCAATGAGGCTTGGGTGGGCGAGGATGGCTGCTGCAACC 300
Qy      241 TATCTTGGCCCTCTATGGCAATGAGGCTTGGGTGGGCGGATGGCTGCTGCTGCCCC 300

Db      301 CGCGGCTTCGGGCTAGTTGGGCCCC 327
Qy      301 CGCGGCTTCGGGCTAGTTGGGCCCC 327

RESULT      5
ID      PCT-US95-10398-118 STANDARD; DNA; UNC; 573 BP.
AC      xxxxxx
DT      01-JAN-1900
DE      Sequence 118, Application PC/TUS9510398.
```

```
CC Sequence 118, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PORCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 118:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: IND8
CC Sequence 573 BP; 97 A; 185 C; 177 G; 114 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db      1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCACAG 60
Qy      1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db      61 GAGCTCAAGTTCCGGCGGTGCGGAGATCGTTGGTGAGTTTACCTGTCGCCGCCAGG 120
Qy      61 GSGGTNNNNNNCCGGTGGCGGTGAGATCGTTGGTGAGTTTACCTGTCGCCGCCAGG 120

Db      121 GCGCCAGGTTGGTGTGCGCGGCGACTAGGAAGACTTCCGAGCGGTGCGCAACCTCGTGA 180
```

```
Qy 121 GCGCCGAGNNGGTGTCGCGCGACTAGGAAGACTTCGAGCGGTCAACAACCTCGTGGC 180
Db 181 AGCGGACACCTATCCCAAGGCTCGCGCGCGAGGTAGGGCTTGGGCTCAGCCCGGG 240
Qy 181 AGCGGACAGCTATCCCAAGGCTCGCGCGCGAGGTAGGGCTTGGGCTCAGCCCGGG 240
Db 241 CACCTTGGCCCTCTATGCAATAGGGCTTGGGCTGGGCGAGGATGGCTCTCTCACCC 300
Qy 241 TATCTTGGCCCTCTATGCAATAGGGCTTGGGCTGGGCGAGTGGCTCTCTCTCCCC 300
Db 301 CGCGGCTCTGGGCTAGTTGGGGCCCC 327
Qy 301 CGCGGCTCTGGGCTAGTTGGGGCCCC 327
```

RESULT 6
ID PCT-US95-10398-115 STANDARD; DNA; UNC; 573 BP.

```
AC xxxxx
DT 01-JAN-1900
DE Sequence 115, Application PC/TUS9510398.
CC Sequence 115, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R. H. AND
CC APPLICANT: PURCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/250/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2076-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 115:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
```

```
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: T10
SQ Sequence 573 BP; 99 A; 181 C; 176 G; 117 T; 0 other;

Query Match      88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATCAGCAGCAATCCTAAACCTCAAGAAAAACCAACGTAACCAACCGCGCCGACAG 60
Qy 1 ATCAGCAGCAATCCTAAACCTCAAGAAAAACCAACGTAACCAACCGCGCCGCTCAK 60

Db 61 GAGGTCAAGTTCGCGCGGGTGGTCAGATCGTTGGTGGAGTTTACCTGTTGCGCGCAGG 120
Qy 61 G:SGTNNNNNNNCGGCTGGCGGTGAGATCGTTGGTGGAGTTTACCTGTTGCGCGCAGG 120

Db 121 GCGCCAGGTTGGGTGTCGCGCGGACTAGCAAGACTTCCGAGCGGTCCGAACCTCGTGA 180
Qy 121 GCGCCAGGTTGGGTGTCGCGCGGACTAGCAAGACTTCCGAGCGGTCCGAACCTCGTGGC 180

Db 181 AGCGCAGCGCTATCCCAAGGCTCGCCAGCCGAGGCTGGGCTCAGGCCCGGG 240
Qy 181 AGCGCAGCGCTATCCCAAGGCTCGCGGCTCGCGGCGAGGCTGGGCTCAGGCCCGGG 240

Db 241 TACCCTTGGCCCTCTATGCAATAGGCGCATGGGCTGGGCGAGGATGGCTCTCTGTCACCC 300
Qy 241 TATCTTGGCCCTCTATGCAATAGGCGCTCGCGGTGGGCGGAGTGGCTCTGTGCCCCC 300

Db 301 CGTGGCTCGCGCGCTAGTTGGGGCCCC 327
Qy 301 CGCGGCTCTCGCGCCCAATTGGGGCCCC 327
```

RESULT 7
ID PCT-US95-10398-111 STANDARD; DNA; UNC; 573 BP.

```
AC xxxxx
DT 01-JAN-1900
DE Sequence 111, Application PC/TUS9510398.
CC Sequence 111, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R. H. AND
CC APPLICANT: PURCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
```

CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 111:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: D1
SQ Sequence 573 BP; 96 A; 184 C; 177 G; 116 T; 0 other;

Query Match 88.6%; Score 281; DB 13; Length 573;
Best Local Similarity 90.2%; Pred. No. 2.29e-180;
Matches 295; Conservative 4; Mismatches 28; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCACAG 60
|||||
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db 61 GAGCTCAAGTTCCGGCGGCTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120
|:|
Qy 61 GGGTNNNNNNCCGGGTGGCGGCTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120

Db 121 GCGCCAGGTTGGTGTGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACCTCGTGC 180
|||||
Qy 121 GCGCCAGGNNGGTGTGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACCTCGTGC 180

Db 181 AGCGGACACCTATCCCAAGCTCGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACCTCGG 240
|||||
Qy 181 AGCGGACAGCTATCCCAAGCTCGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACCTCGG 240

Db 241 TACCCTGGCCCTCTATGGCAAGAGGCTTGGGTGGGCAAGATGGCTCTGTACCC 300
||
Qy 241 TATCTTGGCCCTCTATGGCAATGAGGCTGCGGGTGGCGGNTGGCTCTGTCTCCCC 300

Db 301 CGCGGCTCCCGGCTAGTTGGGGCCCC 327
|||||
Qy 301 CGCGGCTCTCGGCCCAATTGGGGCCCC 327

RESULT 8
ID PCT-US95-10398-112 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 112, Application PC/TUS9510398.
CC Sequence 112, Application PC/TUS9510398

CC GENERAL INFORMATION:
CC APPLICANT: BUKH, J., MILLER, R.H. AND
CC APPLICANT: PURCELL, R.H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 112:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: US6
SQ Sequence 573 BP; 97 A; 183 C; 178 G; 115 T; 0 other;

Query Match 88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGAGCAGCAATCTTAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCACAG 60
|||||
Qy 1 ATGAGCAGCAATCTTAACCTCAAGAAAAACCAACGTAACACCAACCGCGCCCTCAK 60

Db 61 GAGCTCAAGTTCCGGCGGCTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120
|:|
Qy 61 GGGTNNNNNNCCGGGTGGCGGCTGCATCGTTGGTGAGTTTACCTGTTGCCGCGCAGG 120

Db 121 GCGCCAGGTTGGTGTGCGCGGCACTAGGAAGACTTCCGAGCGGTCAACACCTCGTGA 180
|||||

QY	121	GGCCCCAGNNNGGGTGTGCGCGCGACCTAGGAAGACTTCGAGGGGTACAACTCGTGTGGC	180
Db	181	AGCGCAGCAACTTATCCCAAGGCTTCGCGCGCCCGCAGCGCAGGGCCTGGGCTCAGCGCGGG	240
QY	181	AGCGCAGCAGCTATCCCAAGGCTCGYCGGTCGAGGGCAGGTCTTGGGCTCAGCGCGGG	240
Db	241	TACCTTGGCCCCCTTATGGAACGAGGCGCATGGGGTGGGCAAGATGGCTCTGTGTACCC	300
QY	241	TATCTTGGCCCCCTATGGGCAATGAGGCGCTCGGGGTGGCGGGGNTGGCTCTGTCTCCCC	300
Db	301	CGTGGCTCCGGCGCTAGTTGGGGCCCC	327
QY	301	CGCGGCTCTGGGCGCAATTGGGGCCCC	327

RESULT	9
ID	PC7-US95-10398-105 STANDARD; DNA; UNC; 573 BP.
AC	xxxxxx
DT	01-JAN-1900
DE	Sequence 105, Application PC/TUS9510398.
CC	Sequence 105, Application PC/TUS9510398
CC	GENERAL INFORMATION:
CC	APPLICANT: BUKH, J., MILLER, R.H. AND
CC	APPLICANT: PURCELL, R.H.
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC	TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC	NUMBER OF SEQUENCES: 263
CC	CORRESPONDENCE ADDRESS:
CC	

RESULT	10
ID	PCT-US95-10398-124 STANDARD; DNA; UNC; 573 BP.

CC	STRANDEDNESS: single	
CC	TOPOLOGY: linear	
CC	ORIGINAL SOURCE:	
CC	ORGANISM: homosapiens	
CC	INDIVIDUAL ISOLATE: S14	
5Q	Sequence 573 BP; 97 A; 185 C; 176 G; 115 T; 0 other;	
	Query Match 88.0%; Score 279; DB 13; Length 573;	
	Best local Similarity 89.9%; Pred. No. 6.39e-179;	
	Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;	
Db	1 ATGAGCAGCAATCTTAACCTCAAGAAAGAAACCAACGTAACACCAACCGTCGCCACAG 60	
Qy	1 ATGAGCAGCAATCTTAACCTCAAGAAAGAAACCAACGTAACACCAACCGCGCCCTCAK 60	
Db	61 GAGCTCAAGTCTCCGGGTGGGGGTACATGTTGGTGGAGTTTACTTTGTCGGCGCGAG 120	
Qy	61 GGSCTNNNNNNCGGGTGGCGGTACATGTTGGTGGAGTTTACTGTTGCGCGCGCAGG 120	
Db	121 GGGCCTAGATTGGGTGTGGCGCGCAGCGAAGACTTCCGAGCGGTGCGAACCTCGAGGT 180	
Qy	121 GCGCCACAGNNGGTGTGCGCGGACATGGAAGACTTCCGAGCGGTACAACTCTGTGGC 180	
Db	181 AGAGCTCAGCCTATCCCCAAAGCACATCGCGCCCGCAGGACGACTGGGCTCAGCGCGGG 240	
Qy	181 AGGCGCAGCCTATCCCCAAAGGCTGTGCGGTCGAGGCGCAGGTCTGGGCTCAGCCCGGG 240	
Db	241 TATCCTTGGCCCCCTCTATGCAATGAGGCTCGGGGTGGCGGGGATGGCTCTGTCTGCC 300	
Qy	241 TATCCTTGGCCCCCTCTATGGCAATGAGGCTCGGGGTGGCGGGGATGGCTCTGTCTGCC 300	
Db	301 CGTGGCTCTCGGCCCTAGCTGGGGCCCC 327	
Qy	301 CGCGGCTCTCGGCCCAATTCGGGGCCC 327	

RESULT	10	
ID	PC-US95-10398-124 STANDARD; DNA; UNC; 573 BP.	
AC	xxxxxx	
DT	01-JAN-1900	
DE	Sequence 124, Application PC/TUS9510398.	
CC	Sequence 124, Application PC/TUS9510398	
CC	GENERAL INFORMATION:	
CC	APPLICANT: BUKH, J., MILLER, R.H. AND	
CC	APPLICANT: PORCELL, R.H.	
CC	TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED	
CC	TITLE OF INVENTION: AMINO ACID SEQUENCES OF	
CC	TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS	
CC	TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE	
CC	TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES	
CC	NUMBER OF SEQUENCES: 263	
CC	CORRESPONDENCE ADDRESS:	
CC	ADDRESSEE: MORGAN & FINNEGAN	
CC	STREET: 345 PARK AVENUE	
CC	CITY: NEW YORK	
CC	STATE: NEW YORK	
CC	COUNTRY: USA	
CC	ZIP: 10154	
CC	COMPUTER READABLE FORM:	
CC	MEDIUM TYPE: FLOPPY DISK	
CC	COMPUTER: IBM PC COMPATIBLE	
CC	OPERATING SYSTEM: PC-DOS/MS-DOS	
CC	SOFTWARE: WORDPERFECT 5.1	
CC	CURRENT APPLICATION DATA:	

CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIORITY CLAIM DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 124:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homoeapiens
CC INDIVIDUAL ISOLATE: T3
SQ Sequence 573 BP; 98 A; 181 C; 177 G; 117 T; 0 other;


```
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 121:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: HK5
SQ Sequence 573 BP; 103 A; 183 C; 173 G; 114 T; 0 other;

Query Match      88.0%; Score 279; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 6.39e-179;
Matches 294; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGCTCAK 60

Db 61 GAGCTCAAGTTCCCGGGGGTGTGATGCTGTGGTGGAGTTTACCTGTTCGGCGCCGAG 120
Qy 61 GGGTNNNNNNCCGGTGGCGGTGATGCTGTGGTGGAGTTTACCTGTTCGGCGCCGAG 120

Db 121 GCGCCAGAGTTGGTGTGCGCGGACGAGAGACTTCCGAGCGGTTCACACCTCTGGA 180
Qy 121 GCGCCAGAGTTGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTTCACACCTCTGGA 180

Db 181 AGCGCACACCTATCCCAAGGCTCGCGGACGAGGACCTGGGCTCAGCCCGGG 240
Qy 181 AGCGCACAGCTATCCCAAGGCTCGCGGACGAGGACCTGGGCTCAGCCCGGG 240

Db 241 TATCCTTGGCCCTCTATGGCAATGAGGATGGGTTGGGCGAGATGGCTCTGTCAACC 300
Qy 241 TATCCTTGGCCCTCTATGGCAATGAGGCTGGGTTGGGCGGAGTGGCTCTGTCCCCC 300

Db 301 CATGGCTCTCGGCTAGTTGGGCCCC 327
Qy 301 CGCGGCTCTCGGCCAATTGGGCCCC 327

RESULT 14
ID PCT-US95-10398-106 STANDARD; DNA; UNC; 573 BP.
AC xxxxx
DT 01-JAN-1900
DE Sequence 106, Application PC/TUS9510398.
CC Sequence 106, Application PC/TUS9510398
CC GENERAL INFORMATION:
CC APPLICANT: BURKH, J., MILLER, R. H. AND
```

```
CC APPLICANT: PURCELL, R. H.
CC TITLE OF INVENTION: NUCLEOTIDE AND DEDUCED
CC TITLE OF INVENTION: AMINO ACID SEQUENCES OF THE ENVELOPE 1 AND
CC TITLE OF INVENTION: CORE GENES OF ISOLATES OF HEPATITIS C VIRUS
CC TITLE OF INVENTION: AND THE USE OF REAGENTS DERIVED FROM THESE
CC TITLE OF INVENTION: SEQUENCES IN DIAGNOSTIC METHODS AND VACCINES
CC NUMBER OF SEQUENCES: 263
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: MORGAN & FINNEGAN
CC STREET: 345 PARK AVENUE
CC CITY: NEW YORK
CC STATE: NEW YORK
CC COUNTRY: USA
CC ZIP: 10154
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: FLOPPY DISK
CC COMPUTER: IBM PC COMPATIBLE
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: WORDPERFECT 5.1
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/10398
CC FILING DATE: 15-AUG-1995
CC CLASSIFICATION:
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/086,428
CC FILING DATE: 29 JUNE 1993
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: 08/290/665
CC FILING DATE: 15 AUGUST 1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: RICHARD W. BORK
CC REGISTRATION NUMBER: 36,459
CC REFERENCE/DOCKET NUMBER: 2026-4116
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: (212) 758-4800
CC TELEFAX: (212) 751-6849
CC TELEX: 421792
CC INFORMATION FOR SEQ ID NO: 106:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 573 base pairs
CC TYPE: nucleic acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC ORIGINAL SOURCE:
CC ORGANISM: homosapiens
CC INDIVIDUAL ISOLATE: SH1
SQ Sequence 573 BP; 91 A; 182 C; 177 G; 117 T; 0 other;

Query Match      87.7%; Score 278; DB 13; Length 573;
Best Local Similarity 89.9%; Pred. No. 3.37e-178;
Matches 293; Conservative 4; Mismatches 29; Indels 0; Gaps 0;

Db 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGACAG 60
Qy 1 ATGACGACGAATCCTAAACCTCAAGAAACCAACGTAACACCAACCGCGCCGCTCAK 60

Db 61 GAGCTCAAGTTCCCGGGTGGCGGTGATGCTGTGGTGGAGTTTACTTGTTCGGCGCCGAG 120
Qy 61 GGGTNNNNNNCCGGTGGCGGTGATGCTGTGGTGGAGTTTACTTGTTCGGCGCCGAG 120

Db 121 GCGCCAGAGTTGGTGTGCGCGGACGAGGAGACTTCCGAGCGGTTCGCAACCTCTGAG 180
Qy 121 GCGCCAGAGTTGGTGTGCGCGGACTAGGAGACTTCCGAGCGGTTCGCAACCTCTGAG 180
```

WQ5QEL

(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPerch_nn n.a. - n.a. database search, using Smith-Waterman algorithm
Run on: Mon Feb 23 20:52:18 1998; MasPar time 208.41 Seconds
774.811 Million cell updates/sec
Tabular output not generated.

Title: >US-08-836-075A-1
Description: (1-327) from US08836075A.seq
Perfect Score: 317
N.A. Sequence: 1 ATGAGCAGCATCTAAACC.....CTCGGCCCAATGGGGCCCC 327
Comp: TACTCGTGCTTAGGATTGG.....GAGCGGTTAAACCCGGGG

Scoring table:
Gap 6
TABLE default

Mmatch STD : Dbase 0; Query 0

Searched: 665703 seqs, 246912890 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: EST-A
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98
EST-B
99:EST99 100:EST100 101:EST101 102:EST102 103:EST103
104:EST104 105:EST105 106:EST106 107:EST107 108:EST108
109:EST109 110:EST110 111:EST111 112:EST112 113:EST113
114:EST114 115:EST115 116:EST116 117:EST117 118:EST118
119:EST119 120:EST120 121:EST121 122:EST122 123:EST123
124:EST124 125:EST125 126:EST126 127:EST127 128:EST128
129:EST129 130:EST130 131:EST131 132:EST132 133:EST133

134:EST134 135:EST135 136:EST136 137:EST137 138:EST138
139:EST139 140:EST140 141:EST141 142:EST142 143:EST143
144:EST144 145:EST145 146:EST146 147:EST147 148:EST148
149:EST149 150:EST150 151:EST151 152:EST152 153:EST153
154:EST154 155:EST155 156:EST156 157:EST157 158:EST158
159:EST159 160:EST160 161:EST161 162:EST162 163:EST163
164:EST164 165:EST165 166:EST166 167:EST167 168:EST168
169:EST169 170:EST170 171:EST171 172:EST172 173:EST173
174:EST174 175:EST175 176:EST176 177:EST177 178:EST178
179:EST179 180:EST180 181:EST181 182:EST182 183:EST183
184:EST184 185:EST185 186:EST186 187:EST187 188:EST188
189:EST189 190:EST190 191:EST191 192:EST192 193:EST193
194:EST194 195:EST195 196:EST196

Statistics: Mean 9.769; Variance 2.141; scale 4.562

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	No.	Score	Query	Match	Length	DB	ID	Description	Pred. No.
C	1	23	7.3	423	13	T96691	ye52h03.sl	Homo sapie	1.30e-05
	2	22	6.9	301	36	R77276	yi75d01.sl	Homo sapie	1.86e-04
	3	22	6.9	426	92	H79148	yu27g08.sl	Homo sapie	1.86e-04
C	4	22	6.9	491	54	R94705	yq43d01.sl	Homo sapie	1.86e-04
	5	21	6.6	215	28	R49453	yq68e02.sl	Homo sapie	2.43e-03
C	6	21	6.6	217	85	N56240	J9576F	Homo sapiens c	2.43e-03
C	7	21	6.6	227	55	R97434	yq53h06.sl	Homo sapie	2.43e-03
C	8	21	6.6	304	14	R07673	ye98c07.rl	Homo sapie	2.43e-03
	9	21	6.6	332	140	AA038712	mi92c09.rl	Soares mou	2.43e-03
C	10	21	6.6	404	147	AA119984	mn34a10.rl	Beddington	2.43e-03
C	11	21	6.6	419	92	H81577	yu73b06.sl	Homo sapie	2.43e-03
	12	21	6.6	459	37	R82259	yj17g01.rl	Homo sapie	2.43e-03
	13	21	6.6	468	183	N53148	yv55q10.sl	Soares fet	2.43e-03
	14	20	6.3	210	35	R76212	yi71d11.sl	Homo sapie	2.89e-02
C	15	20	6.3	276	68	H40395	yp60e05.rl	Homo sapie	2.89e-02
	16	20	6.3	305	59	T31232	EST28924	Homo sapiens	2.89e-02
	17	20	6.3	370	95	ATTS3506	A. thaliana	transcrib	2.89e-02
	18	20	6.3	374	178	AA192539	zp91g05.rl	Stratagene	2.89e-02
	19	20	6.3	379	29	R33992	yj70c03.sl	Homo sapie	2.89e-02
C	20	20	6.3	399	39	H13389	yj72b11.rl	Homo sapie	2.89e-02
	21	20	6.3	402	16	T09138	EST07031	Homo sapiens	2.89e-02
C	22	20	6.3	415	45	H17241	ym42f08.rl	Homo sapie	2.89e-02
	23	20	6.3	418	2	T59290	yc02c12.rl	Homo sapie	2.89e-02
	24	20	6.3	437	47	H50371	yo29b01.rl	Homo sapie	2.89e-02
C	25	20	6.3	437	45	H17422	ym40e09.rl	Homo sapie	2.89e-02
	26	20	6.3	442	26	R42917	yq14b10.sl	Homo sapie	2.89e-02
C	27	20	6.3	443	38	H09205	y198d07.rl	Homo sapie	2.89e-02
	28	20	6.3	445	92	H79969	yv79f12.rl	Homo sapie	2.89e-02
C	29	20	6.3	474	56	H19576	yn59c01.rl	Homo sapie	2.89e-02
	30	20	6.3	478	51	R85141	yo41d07.rl	Homo sapie	2.89e-02
C	31	20	6.3	482	121	N93628	zb70e11.sl	Soares fet	2.89e-02
	32	20	6.3	514	138	AA028338	mi20f11.rl	Soares mou	2.89e-02
	33	20	6.3	514	46	H46973	yo19g12.rl	Homo sapie	2.89e-02
	34	19	6.0	81	170	AA155886	zo70c12.rl	Stratagene	3.10e-01
	35	19	6.0	331	93	H84640	ye67e06.sl	Homo sapie	3.10e-01
	36	19	6.0	338	175	AA181878	zp66a02.sl	Stratagene	3.10e-01
	37	19	6.0	345	180	AA022461	ze70a03.sl	Soares fet	3.10e-01
	38	19	6.0	391	109	W53431	md44d04.rl	Soares mou	3.10e-01
	39	19	6.0	422	158	W81106	zd82b02.sl	Soares fet	3.10e-01

Molecular Cardiology
University of Toronto
Banning Institute, 100 College St., Toronto, Ontario, M5G1L5
Tel: 416/9788758
Fax: 416/9785650
Email: liewc@utcc.utoronto.ca.

FEATURES	Location/Qualifiers
source	1..217

ORIGIN	BASE COUNT	mRNA	<1..>217	40 a	62 c	69 g	46 t
1	1	U	U	U	U	U	U
2	2	U	U	U	U	U	U
3	3	U	U	U	U	U	U
4	4	U	U	U	U	U	U
5	5	U	U	U	U	U	U
6	6	U	U	U	U	U	U
7	7	U	U	U	U	U	U
8	8	U	U	U	U	U	U
9	9	U	U	U	U	U	U
10	10	U	U	U	U	U	U
11	11	U	U	U	U	U	U
12	12	U	U	U	U	U	U
13	13	U	U	U	U	U	U
14	14	U	U	U	U	U	U
15	15	U	U	U	U	U	U
16	16	U	U	U	U	U	U
17	17	U	U	U	U	U	U
18	18	U	U	U	U	U	U
19	19	U	U	U	U	U	U
20	20	U	U	U	U	U	U
21	21	U	U	U	U	U	U
22	22	U	U	U	U	U	U
23	23	U	U	U	U	U	U
24	24	U	U	U	U	U	U
25	25	U	U	U	U	U	U
26	26	U	U	U	U	U	U
27	27	U	U	U	U	U	U
28	28	U	U	U	U	U	U
29	29	U	U	U	U	U	U
30	30	U	U	U	U	U	U
31	31	U	U	U	U	U	U
32	32	U	U	U	U	U	U
33	33	U	U	U	U	U	U
34	34	U	U	U	U	U	U
35	35	U	U	U	U	U	U
36	36	U	U	U	U	U	U
37	37	U	U	U	U	U	U
38	38	U	U	U	U	U	U
39	39	U	U	U	U	U	U
40	40	U	U	U	U	U	U
41	41	U	U	U	U	U	U
42	42	U	U	U	U	U	U
43	43	U	U	U	U	U	U
44	44	U	U	U	U	U	U
45	45	U	U	U	U	U	U
46	46	U	U	U	U	U	U
47	47	U	U	U	U	U	U
48	48	U	U	U	U	U	U
49	49	U	U	U	U	U	U
50	50	U	U	U	U	U	U
51	51	U	U	U	U	U	U
52	52	U	U	U	U	U	U
53	53	U	U	U	U	U	U
54	54	U	U	U	U	U	U
55	55	U	U	U	U	U	U
56	56	U	U	U	U	U	U
57	57	U	U	U	U	U	U
58	58	U	U	U	U	U	U
59	59	U	U	U	U	U	U
60	60	U	U	U	U	U	U
61	61	U	U	U	U	U	U
62	62	U	U	U	U	U</	

Query Match 6.6%; Score 21; DB 85; Length 217;
Best Local Similarity 61.9%; Pred. No. 2.43e-03;
Matches 26; Conservative 2; Mismatches 14; Indels 0; Gaps 0;

This clone is available royalty-free through LINL ; contact the
 IMAGE Consortium (info@image.lnl.gov) for further information.

DEFINITION Yq53h06.s1 Homo sapiens cDNA clone 195547 3'.

ACCESSION R97434

NID g983094

KEYWORDS EST.

SOURCE

human clone-199547 primer=Promega -2m13 library=Soares fetal liver spleen INFLS vector=p773D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rstel= Pac I Rsite2=Eco RI liver and spleen from a 20 week post conception male fetus. 1st strand cDNA was primed with a Pac I - oligo(dT) primer [5' AACTCGAAGAAATTAAGAATCTTTTTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac I and cloned into the Pac I and Eco RI sites of the modified p773 vector. Library went through one round of normalization. Library constructed by Bento Soares and M.Fátima Bonaldo.

ORGANISM

Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Archonta; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 227)

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.

TITLE
JOURNAL
The WashU-Merck EST Project
Unpublished (1995)

Homosapiens
Homo sapiens
Eumetazoa; Bilateria; Coelomata;
Eukaryotae; Metazoa; Eumetazoa;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.
1 (bases 1 to 217)

Email: est@watson.wustl.edu
 High quality sequence stops: 201
 Source: IMAGE Consortium, L1NL
 This clone is available royalty--
 IMAGE Consortium (info@image.llnwd.net)

Contact: Liew CC

FEATURES

source

Location/Qualifiers

1..227

/organism="Homo sapiens"

/clone="199547"

<1..>227

BASE COUNT 50 a 57 c 76 g 42 t 2 others

ORIGIN

Query Match 6.6%; Score 21; DB 55; Length 227;

Best Local Similarity 68.2%; Pred. No. 2.43e-03;

Matches 30; Conservative 2; Mismatches 12; Indels 0; Gaps 0;

Db 152 cgggctgagaccctcctgctgctgagccagcaggtaggg 195

||||||| || | ||||| : | : | | | | |

Cp 238 CGGGTGAGCCGAGGACCTGCTCGGCCGCGAGCCTGGGG 195

RESULT 8

LOCUS R07673 304 bp mRNA EST 05-APR-1995

DEFINITION ye98c07.r1 Homo sapiens cDNA clone 125772 5' similar to contains

Alu repetitive element;

ACCESSION R07673

NID g759596

KEYWORDS EST.

SOURCE

human clone=125772 library=Soares fetal liver spleen INTLS
vector=pT73D (Pharmacia) with a modified polylinker host=DH10B
(ampicillin resistant) primer=M13Rpl Reitel=Pac I Reitel=Eco RI
Liver and spleen from a 20 week-post conception male fetus. 1st
strand cDNA was primed with a Pac I - oligo (dT) primer [5'
AACTGAGAAATTAATTAAGATCTTTTTTTTTTTTTTTT 3'], double-stranded
cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Pac
I and cloned into the Pac I and Eco RI sites of the modified pT73
vector. Library went through one round of normalization. Library
constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens

Eucaryotae; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;

Eutheria; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 304)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, E., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

High quality sequence stops: 265

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

FEATURES

source

Location/Qualifiers

1..304

/organism="Homo sapiens"

/clone="125772"

BASE COUNT 80 a 78 c 53 g 91 t 2 others

ORIGIN

Query Match 6.6%; Score 21; DB 14; Length 304;

Best Local Similarity 85.7%; Pred. No. 2.43e-03;

Matches 24; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 102 gggacatgagccaccagccagccagcgca 129

||||| ||||| | ||||| |||||

Cp 298 GGCACAGGAGCCAGCCGCCACCCGCA 271

RESULT 9

LOCUS AA038712 332 bp mRNA EST 28-AUG-1996

DEFINITION m192c09.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 474064 5'

similar to gb:X53476 Mouse mRNA for non-histone chromosomal protein

HMG-14 (MOUSE);

ACCESSION AA038712

NID g1514121

KEYWORDS EST.

SOURCE house mouse.

ORGANISM

Mus musculus

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;

Murinae; Mus.

1 (bases 1 to 332)

Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,

Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,

Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and

Waterston, R.

The WashU-HMI Mouse EST Project

Unpublished (1996)

JOURNAL

COMMENT

Contact: Marra M/Mouse EST Project

WashU-HMI Mouse EST Project

Washington University School of MedicineP

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: mouseest@watson.wustl.edu

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

MGI:284808

Seq primer: -28M13 rev2 from Amersham

High quality sequence stop: 187.

FEATURES

source

Location/Qualifiers

1..332

/organism="Mus musculus"

/note="Vector: pT73D (Pharmacia) with a modified

polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

was primed with a Not I - oligo(dT) primer [5'

TGTTACCAATCTGAAGTCGAGCGCGCATTTTTTTTTTTTTTTT 3'] ,

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into

the Not I and Eco RI sites of a modified pT73 vector

(Pharmacia). Library went through one round of

normalization to a Cot = 5. Library constructed by Bento

Soares and M.Fatima Bonaldo. RNA was kindly provided by

Dr. Minoru Ko (Wayne State University)."

/clone="474064"

/dev_stage="19.5 dpc total fetus"

/lab_host="DH10B (ampicillin resistant)"

<1..>332

83 a 56 c 92 g 101 t

BASE COUNT

ORIGIN

```

Query Match      6.6%; Score 21; DB 140; Length 332;
Best Local Similarity 92.0%; Pred. No. 2, 43e-03;
Matches 23; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

RESULT	10								
LOCUS		A1119984	404 bp	mRNA	EST	19-NOV-1996			
DEFINITION		mm34a10.r1	Bedlington mouse	embryonic region	Mus musculus	cDNA			
		clone 539802	5'	similar to	FR:G1236085	G1236085	L1SCH7	;	
ACCESSION		A1119984							
NID		q1677684							
KEYWORDS		EST.							
SOURCE		house mouse.							
ORGANISM		Mus musculus							
		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;							
		Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;							
		Mus.							

1 (bases 1 to 404)

REFERENCE

Marras, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Gelsel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)

Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LINL ; contact the
IMAGE Consortium (info@image.linn.gov) for further information.
MGI:326738

Putative full length read
vector to vector length is 425
Seq primer: -40m13 ET
High quality sequence stop: 362.

```

FEATURES
source
Location/Qualifiers
1..404
/organism="Mus musculus"
/strain="C57BL6 x DBA"
/note="Vector: pCMV-SPORT; Site 1: SalI; Site 2: NotI;
Cloned unidirectionally. Primer: Oligo dT. Gastrulating
embryos were collected at 7.5dpc from C57BL6 x DBA
matings, excluding embryos that had developed head folds
and all extraembryonic tissues. Average insert size: 1.3
kb (range: 0.5 - 3.0 kb). Referenced in Development 121,
2479-2489 (1995)."
/clone="539802"
/clone_lib="Beddington mouse embryonic region"
/seq="rooted"

```

mRNA	BASE COUNT	ORIGIN
<1..>404	107 a 94 c	109 g 94 t

	Query Match	Best Local Similarity	Score 21; DB 147;	Length 404;
	Matches 38; Conservative 1;	66.7%;	Pred. No. 2.43e-03;	
	Mismatches 18; Indels 0; Gaps 0;			
Db	44	ggctaggagtcacacgtgctgcacccagatcccgctcagcgatccacagatccctcg	100	
Cp	252	ggccacagataccgcggctgcacgacgactgccttcggatccgcgcgcgcgccttcgg	196	

RESULT	11				
LOCUS	H81577	419 bp	mRNA	EST	
DEFINITION	Y073806.s1	Homo sapiens cDNA clone 239411 3'.			
ACCESSION	H81577				
	NID	g1114654			
KEYWORDS	EST.				

SOURCE human clone-239411 primer=Promega-2ml3 library=Weizmann Olfactory Epithelium vector=pBluescript SK- host=SO18 cells (kanamycin resistant) Restr-EcoRI SmaEcoI XhoI from 35 year old female. The cDNA was oligo (dT) primed with an XhoI restriction enzyme recognition site and an 18 base poly dT sequence. For the 5' end, the synthesized cDNA termini were treated with T4 DNA polymerase and EcoRI adaptors were ligated to the blunt ends.

ORGANISM Homo sapiens
Eukaryotae; Metazoa; Eumetazoa; Bilateria; Coelomata;
Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes;
Sarcopterygii; Choanata; Tetrapoda; Amniota; Mammalia; Theria;
Eutheria; Archonta; Primates; Catarrhini; Hominoidea; Homo.

1 (bases 1 to 419).

REFERENCE
AUTHORS
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, R., Hulman, M., Kucaba, T., Le, N., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaske, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.
TITLE
WashU-Merck EST Project
Unpublished (1995)

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
High quality sequence stops: 294
Source: IMAGE Consortium, LINL
This clone is available royalty-free through LINL / contact the
IMAGE Consortium (info@image.linl.gov) for further information.

FEATURES	source	Location/Qualifiers
	1..419	
	/organism="Homo sapiens"	
	/clone="239411"	
	<1..419	
mRNA		
BASE COUNT	121 a 70 c 62 g 164 t	2 others
ORIGIN		
Query Match	6.6%;	Score 21; DB 92; Length 419;
Best Local Similarity	61.5%;	Pred. No. 2.43e-03;
Matches	40: Conservative	2: Mismatches 22: Indels
		1: Gaps 1:

Db 46 ggaatgcaaacctcatgaggcgaggttttgatttattgtattgtttacaccaggatt 105
|| || :||||| | || ||| ||| | |||| | ||| ||
Cn 74 GGNNNNNAWCWCMGTAGGCGGGCGGT-GCTTACGTTTGTTTTCTTTGAGGTTT 16

Db 106 aqaat 110

LOCUS	AA282971	226 bp	mRNA	EST	15-MAY-1997
DEFINITION	ztl4a08.s1	Soares	NBHTCBC	Homo sapiens cDNA clone 713078 3'.	
ACCESSION	AA282971				
NID	g1925885				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;				
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;				
	Homo.				
REFERENCE	1 (bases 1 to 226)				
AUTHORS	Hillier, L., Allen, M., Bowles, L., Dubuque, T., Geisel, G., Jost, S., Kucaba, T., Lacy, M., Le, N., Lennon, C., Marra, M., Martin, J., Moore, B., Schellenberg, K., Steptoe, L., Tan, F., Theising, B., White, Y., Wylie, T., Waterston, R. and Wilson, R.				
TITLE	WASHU-Merck EST Project 1997				
JOURNAL	Unpublished (1997)				

NTID	g1135707		
KEYWORDS	EST.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	1 (bases 1 to 326)		
AUTHORS	Landfanchi,G., Muraro,T., Caldars,F., Pacchioni,B., Pallavicini,A., Pandolfo,D., Toppo,S., Trevisan,S., Scarso,S. and Valle,G.		
TITLE	Identification of 4,370 expressed sequence tags (ESTs) from a 3'-end specific cDNA library of human skeletal muscle by DNA sequencing and filter hybridisation		
JOURNAL	Genome Res. 1, 35-42 (1996)		
REFERENCE	2 (bases 1 to 326)		
AUTHORS	Valle,G.		
TITLE	Direct Submission		
JOURNAL	Submitted (19-DEC-1995) CRIBI Biotechnology Centre, University of Padua, Via Trieste 75, 35121 Padua, Italy		
COMMENT	ABI Chromatograms and other information are available on WWW at http://eos.bio.unipd.it		
	Vector: pcDNAII (Invitrogen)		
	V type: Plasmid		
	RE_1: EcoRI		
	RE_2: NotI		
	PRIMER: PC2R		
	DESCR: The cDNA was constructed using a biotin-NotI-oligo(dT) primer, sonicated and size selected (450-550 BP). The biotinylated 3'-ends were affinity purified, directionally cloned and sequenced (5' -> 3').		

[illegible]


```

MG1:321443
Seq primer: -28ml3 revl ET from Amersham.
          Location/Qualifiers
            1..389
              /organism="Mus musculus"
              /note="Vector: pBluescript SK-; Site 1: EcoRI; Site 2: XhoI; Cloned unidirectionally.. Primer: Oligo dTm. P19 cell line treated with retinoic acid. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCACCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'."
              /clone="534507"
              /clone.lib="Stratagene mouse embryonic carcinomara (#937318)".
              /dev stage="RA-treated P19 cell line (ATCC CRL1825)"
              /lab host="SOLR (kanamycin resistant)"
<1..>389
      mRNA                      98 a   105 c    117 g     69 t
EASE COUNT
ORIGIN
Query Match           6.6%; Score 21; DB 9; Length 389;
Best Local Similarity 66.7%; Pred. No. 1.20e-03;
Matches 38; Conservative 1; Mismatches 18; Indels 0; Gaps 0;

Db      204   ggctaggatccaagtgcctgaccacagatcccgtcaytgattccacgatcctcggg 260
        || | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Cd      252   ggcgccaaagataccggcgctgagccacagaccttgccttcggrccgcagcgtttggg 196

```

	8								
RESULT	A4437253	421 bp	mRNA	EST	30-MAY-1997				
LOCUS	zv54608.s1	Soares testis NHT Homo sapiens cDNA clone 757478 3'.							
DEFINITION	A4437253	NID							
ACCESSION	Q2142167	EST.							
KEYWORDS		human.							
SOURCE	Homo sapiens	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;							
ORGANISM	Vertebrates; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;								
	Rommo.								
REFERENCE	1 (bases 1 to 421)								
AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubouque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Ie,N., Lennon,C., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wyllie,T., Waterston,R. and Wilson,R.								
TITLE	WashU-Merck EST Project 1997								
JOURNAL	Unpublished (1997)								

Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL ; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Seq primer: -41ml3 fwd. ET from Amersham
High quality sequence stop: 291.
Location/Qualifiers

```

1.1.421
source
/organism="Homo sapiens"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA

```

was prepared from mRNA obtained from Clontech Laboratories, Inc., and primed with a Not I - oligo(dT) primer [5' TCTTACCAATCTCAAGTGGAGCGGCGCCGCAATTTTTTTTTT 3'].

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization to Cot5, and was constructed by Bento Soares and M. Fatima Bonaldo."

/clone="757478"
 /sex="male"
 /lab_host="DH10B"
 complement(<1..>421)
 /db_xref="GDB:5978092"

75 a 122 c 151 g 73 t

BASE COUNT
 ORIGIN

Query Match	6.6%;	Score 21;	DB 89;	Length 421;
Best Local Similarity	82.8%;	Pred. No. 1.20e-03;		
Matches 24;	Conservative 1;	Mismatches 4;	Indels 0;	Gaps 0;
Db	377	gctggggcggagggcacgtctctgggcacag	405	
Ov	206	cygggycggagggcgaggtctctgggctcag	234	

RESULT	9								
LOCUS	AA126152	427 bp	mRNA	EST	19-MAY-1997				
DEFINITION	z18sh07.r1 Stratagene colon (#937204) Homo sapiens cDNA clone								
	511453 5' similar to FR:G451854 G451854 ALPHA-MANNOSIDASE ; .								
ACCESSION	AA126152								
NID	gl685800								
KEYWORDS	EST.								
SOURCE	human.								
ORGANISM	Homo sapiens								
	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;								
	Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;								
	Homo.								
REFERENCE	1 (bases 1 to 427)								
AUTHORS	Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawking, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marr, M., Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.								
TITLE	WashU-Merck EST Project								
COMMENT	Unpublished (1995)								

Contact: **Wilson RK**
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Insert Length: 2041 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 240.

```

FEATURES
source
Location/Qualifiers
1..427
/organism="Homo sapiens"
/note="Organ: colon; Vector: pBluescript SK-; Site 1:
EcoRI; Site 2: XhoI; Cloned unidirectionally. Primer:

```

Oligo dT. T-84 colonic epithelial cell line. Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' GAATTCGGCAGCAG 3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'.

/clones="511453"

/clone_lib="Stratagene colon (#937204)"

/lab_host="SOLR cells (kanamycin resistant)"

<1..>427

/db_xref="GDB:3843948"

BASE COUNT 74 a 125 c 147 g 76 t 5 others

ORIGIN

Query Match

Best Local Similarity 6.6%; Score 21; DB 94; Length 427;

Matches 20; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 219 ccctcagtggtgctgcccgcgggtggcg 248

|||||: ||| |||||||||

Qy 54 CCTCAGCGCTNNNNNNCCGGTGGCGG 83

RESULT 10

LOCUS AA410657 450 bp mRNA EST 18-MAY-1997

DEFINITION zt29g04.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 723798 5'.

ACCESSION AA410657

NID g2069762

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

REFERENCE 1 (bases 1 to 450)

AUTHORS

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 443.

FEATURES

source

1..450

/organism="Homo sapiens"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5'

TCGTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaudo."

/clone_lib="Soares ovary tumor NbHOT"

/clone="723798"

/clone_lib="Soares ovary tumor NbHOT"

/sex="Female"

/tissue_type="ovarian tumor"

/lab_host="DH10B (ampicillin resistant)"

<1..>450

/db_xref="GDB:5935183"

BASE COUNT 86 a 135 c 144 g 83 t 2 others

ORIGIN

Query Match

Best Local Similarity 82.8%; Pred. No. 1.20e-03;

Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 273 cctgtccaggagctgcctcgcccgccagc 301

|||||: ||| |||||||||

Cp 234 CTGAGCCGAGGACCTGCCCTCGGCGCGC 206

RESULT 11

LOCUS AA402119 473 bp mRNA EST 16-MAY-1997

DEFINITION zu55404.r1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 741895 5' similar to FR:G397579 G397579 L15 MRNA. ;.

ACCESSION AA402119

NID g2056110

KEYWORDS EST.

SOURCE human.

ORGANISM

Homo sapiens

Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;

Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae;

Homo.

REFERENCE 1 (bases 1 to 473)

AUTHORS

Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.

WashU-Merck EST Project 1997

Unpublished (1997)

COMMENT

Contact: Wilson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

This clone is available royalty-free through LML; contact the IMAGE Consortium (info@image.llnl.gov) for further information.

Seq primer: -28ml3 rev2 ET from Amersham

High quality sequence stop: 462.

FEATURES

source

1..473

/organism="Homo sapiens"

/note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5' TCTTACCAATCTGAAGTGGAGCGCGGTTTTTTTTTTTTTTT 3'],

double-stranded cDNA was size selected, ligated to Eco RI

adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT7T3 vector

(Pharmacia). Library constructed by Bento Soares and

M.Fatima Bonaudo."

/clone="741895"

/clone_lib="Soares ovary tumor NbHOT"

Feb 23 20:55

homefuller/feb1098/US-08-836-075A-1.rstb

13

```

/dev stage="female"
/tissue type="ovarian tumor"
/lab host="DH10B (ampicillin resistant)"
<1..>473
/db xref="GDB:5942137"
BASE COUNT 88 a 153 c 147 g 85 t
ORIGIN

Query Match 6.6%; Score 21; DB 79; Length 473;
Best Local Similarity 82.8%; Pred. No. 1.20e-03;
Matches 24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Db 260 ctgtgccaggagctgcccctgccgcagc 288
||||| ||||||| ||||||| ||| :|
Cp 234 CTGAGCCGAGGACCTGCCCTCGGCGCGC 206

RESULT 12
LOCUS AA031467 216 bp mRNA EST 09-MAY-1997
DEFINITION zkl605.r1 Soares pregnant uterus NbHPU Homo sapiens cDNA clone
470696 5'.
ACCESSION AA031467
NID g1501533
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Primates; Catarrhini; Hominoidea;
Homo.
REFERENCE 1 (bases 1 to 216)
AUTHORS Hillier, L., Clark, M., Dubuque, T., Elliston, K., Hawkins, M.,
Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,
Parsons, J., Rifkin, L., Rohlfing, T., Tan, F., Trevaskis, E.,
Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
TITLE WashU-Merck EST Project
JOURNAL Unpublished (1995)
COMMENT
Contact: Wilson RK
WashU-Merck EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 936 Std Error: 0.00
Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 188.
Location/Qualifiers
1..216
/organism="Homo sapiens"
/Note="Organ: uterus; Vector: pT7T3-Pac; Site 1: Not I;
Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
oligo(dT) primer [5'
AACTGGAAGAAATTCGGCGCGCTTTTTTTTTTTTTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library
went through one round of normalization. Library
constructed by M. Fatima Bonaldo."
/clone="470696"
/clone lib="Soares pregnant uterus NbHPU"
/sex="female"

```

Feb 23 20:55

homefuller/feb1098/US-08-836-075A-1.rstb

14

```

/dev stage="adult"
/lab host="DH10B"
<1..>216
/db xref="GDB:3756698"
BASE COUNT 41 a 92 c 52 g 29 t 2 others
ORIGIN

Query Match 6.3%; Score 20; DB 61; Length 216;
Best Local Similarity 72.1%; Pred. No. 1.44e-02;
Matches 31; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

Db 22 gccggccagtgaggagcagcagccagccaccacccccc 64
||||| ||||||| ||||||| ||| ||||||| ||
Cp 313 GCCGAGCGCGCGGGGAGCAGGAGCCANCCGCCACCGCGCA 271

RESULT 13
LOCUS AA396794 259 bp mRNA EST 25-APR-1997
DEFINITION vb28h10.r1 Soares mouse lymph node NbMLN Mus musculus cDNA clone
750307 5'.
ACCESSION AA396794
NID g2049801
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae;
Murinae; Mus.
REFERENCE 1 (bases 1 to 259)
AUTHORS Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,
Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,
Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,
Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and
Waterston, R.
TITLE The WashU-HMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT
Contact: Marra M/Mouse EST Project
WashU-HMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MG1:459291
Putative full length read
vector to vector length is 260
Seq primer: -28m13 rev2 ET from Amersham.
Location/Qualifiers
1..259
/organism="Mus musculus"
/strain="C57BL/6J"
/Note="Vector: pT7T3-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTCACATCTGAAGTGGAGCGCGGATCTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. RNA
provided by Dr. Bertrand Jordan. Library constructed and
normalized by Bento Soares and M.Fatima Bonaldo."
/clone="750307"

```


Query Match 6.3%; Score 20; DB 66; Length 312;
Best Local Similarity 72.2%; Pred. No. 1.44e-02;
Matches 26; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

Db 135 cccaaggttggccgtctgggtgcaggcctgtgct 170
||||| |:|:| | |||| |||| ||
QY 196 CCAAGGCTCGYCGGYCGAGGCGAGTCTCGGCT 231

Search completed: Mon Feb 23 20:58:41 1998
Job time : 135 secs.

[illegible]

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

M2srch nn n.a. - n.a. database search, using Smith-Waterman algorithm

```
Run on: Mon Feb 23 20:59:06 1998; MasPar time 122.41 Seconds
        673.945 Million cell updates/sec
```

Tabular output not generated.

```

Title:
Description:
Perfect Score:
N.A. Sequence:
Comp:

```

Scoring table: TABLE default
Gap 6

Nmatch STD : Dbase 0: Overv 0

Searched: 333433 seqs. 126143548 bases x 2

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database:

STS

1:STS1 2:STS2 3:STS2 4:STS4 5:STS5 6:STS6 7:STS7 8:STS8
9:STS9 10:SPS10 11:STS11 12:SPS12 13:STS13 14:qNEST1
15:qNEST2 16:qNEST3 17:qNEST4 18:qNEST5 19:qNEST6
20:qNEST7 21:qNEST8 22:qNEST9 23:qNEST10 24:qNEST11
25:qNEST12 26:qNEST13 27:qNEST14 28:qNEST15 29:qNEST16
30:qNEST17 31:qNEST18 32:qNEST19 33:qNEST20 34:qNEST21
35:qNEST22 36:qNEST23 37:qNEST24 38:qNEST25 39:qNEST26
40:qNEST27 41:qNEST28 42:qNEST29 43:qNEST30 44:qNEST31
45:gnSTS 46:enEST1 47:enEST2 48:enEST3 49:enEST4
50:enEST5 51:enEST6 52:enEST7 53:enEST8 54:enEST9
55:enEST10 56:enEST11 57:enEST12 58:enEST13 59:enEST14
60:enEST15 61:enEST16 62:enEST17 63:enEST18 64:enEST19
65:enEST20 66:enEST21 67:enEST22 68:enEST23 69:enEST24
70:enEST25 71:enEST26 72:enEST27 73:enEST28 74:enEST29
75:enEST30 76:enEST31 77:enEST32 78:enEST33 79:enEST34
80:enEST35 81:enEST36 82:enEST37 83:enEST38 84:enEST39
85:enEST40 86:enEST41 87:enEST42 88:enEST43 89:enEST44
90:enEST45 91:enEST46 92:enEST47 93:enEST48 94:enSTS
95:enEST1 96:enEST2

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	22	6,9	156	1	D33607S	D. melanogaster STS d
2	22	6,9	226	52	AA12871494	zt14a08.s1 NCI CGAP_G
3	22	6,9	226	52	AA282971	zt14a08.s1 NCI CGAP_G
4	22	6,9	350	27	AA480207	ne30h02.s1 NCI CGAP_C
5	22	6,9	350	27	AA480207	ne30h02.s1 NCI CGAP_C
c 6	22	6,9	458	78	HSAA1271527	zx77b02.s1 Soates ova
c 7	21	6,6	294	8	G25421	human STS EST405749.
c 8	21	6,6	341	79	HSAA64094	zx82h04.r1 Soates ova
c 9	21	6,6	341	79	AA464094	zx82h04.r1 Soates ova
c 10	21	6,6	419	9	G29737	human STS SHGC-35139.
11	21	6,6	421	59	HS1238538	zv34e08.s1 Soates tes
12	21	6,6	427	76	HSAA26168	zt185h07.r1 Stratagene
c 13	21	6,6	450	54	HS1207341	zt29g04.r1 Soates ova
c 14	21	6,6	475	64	HS1201862	zv35d04.r1 Soates ova
c 15	20	6,3	273	59	HS1300003	aa60c02.r1 NCI CGAP_G
16	20	6,3	275	35	AA504541	aa60c02.r1 NCI CGAP_G
17	20	6,3	291	95	MM1137730	7b073F10 Chromosome 7
18	20	6,3	358	96	MM1139801	mv35g01.r1 Soates mou
c 19	20	6,3	378	57	HS126722	zv48c12.r1 Soates ova
20	20	6,3	382	77	HSAA47251	zt150d06.r1 Soates pre
21	20	6,3	393	72	HS1316949	ni83e09.s1 NCI CGAP_P
22	20	6,3	393	42	AA525270	ni83e09.s1 NCI CGAP_P
23	20	6,3	395	85	MM1295319	vi86f10.r1 Stratagene
24	20	6,3	395	34	AA498868	vi86f10.r1 Stratagene
25	20	6,3	402	91	T09138	EST07031 Homo sapiens
c 26	20	6,3	435	51	HS1182879	zs78h01.r1 NCI CGAP_G
c 27	20	6,3	435	51	AA278629	zs78h01.r1 NCI CGAP_G
c 28	20	6,3	453	76	HSNA38042	zr97b06.r1 NCI CGAP_G
c 29	20	6,3	453	14	AA215509	zr97b06.r1 NCI CGAP_G
c 30	20	6,3	473	77	HSAA47482	zls0a06.r1 Soates pre
c 31	20	6,3	546	13	HUMUT27	Human STS UT27.
32	19	6,0	146	22	AA282835	zt15f03.r1 NCI CGAP_G
33	19	6,0	146	52	HS1187378	zt15f03.r1 NCI CGAP_G
c 34	19	6,0	306	95	HSNA78095	7P01E12 Chromosome 7
c 35	19	6,0	306	11	HUMSWS3587	human chromosome 7 ST
c 36	19	6,0	302	16	HUMSWS3587	human chromosome 7 ST
c 37	19	6,0	355	22	AA281664	zt06g04.r1 NCI CGAP_G
c 38	19	6,0	370	75	HSR083	EST05698 Homo sapiens
c 39	19	6,0	380	80	HSZ2254589	EST56223 Infant brain
c 40	19	6,0	410	54	HS1201751	zu53b11.s1 Soates ova
c 41	19	6,0	460	56	HS1223482	zv38b04.s1 Soates tes
c 42	19	6,0	486	25	AA475558	vh13g10.r1 Soates mou
c 43	19	6,0	515	47	DMAA91246	LD10447.5prime LD Dro
44	19	6,0	531	3	G06885	human STS r1-B8355.
45	19	6,0	556	81	MM1243815	v6f1d1.r1 Birstead M

ALIGNMENTS

RESULT	1
LOCUS	DM36D7S 156 bp DNA STS
DEFINITION	D. melanogaster STS determined from European Mapping Project cosmid.
ACCESSION	Z70980
NID	g1263754
KEYWORDS	sequence tagged site.
SOURCE	fruit fly.

```
ORGANISM      Drosophila melanogaster
Eukaryotae; mitochondrial eukaryotes; Metazoa; Arthropoda;
Tracheata; Insecta; Pterygota; Diptera; Brachycera; Muscomorpha;
REFERENCE     1 (bases 1 to 156)
AUTHORS       European Drosophila Mapping Consortium.
TITLE         Direct Submission
JOURNAL       Submitted (15-APR-1996) Michael Ashburner, Department of Genetics,
              Downing St., Cambridge CB2 3EH, England
COMMENT       STS_name = Dm36D75
              clone_name = 36D7
              STS from promoter = SP6
              vector_class = cosmid, Loxist 6
              origin_of_clone = Oregon-R
              in_situ_site_primary = 98C
              BLAST_program = BLASTN
              database_searched = EMBL
              database_version = 45.0 and updates till date_of_search
              date_of_search = 08-01-1996
              BLAST_program = BLASTX
              database_searched = SWISSPROT
              database_version = 32.0
              date_of_search = 15-12-1995.
              Location/Qualifiers
                source
                  1..156
                    /organism="Drosophila melanogaster"
                    /strain="Oregon-R"
                    /clones="36D7"
BASE COUNT    28 a 37 c 30 g 27 t 34 others
ORIGIN
Query Match   6.9%; Score 22; DB 1; Length 156;
Best Local Similarity 19.4%; Pred. No. 1.48e-04;
Matches 6; Conservative 20; Mismatches 5; Indels 0; Gaps 0;

Db 29 rrcgkrirgagksksrgrgkrksrgrgs 59
      ::::::| :::::| :::::| :::::| :::::|
Qy 259 GGCATGAGGGCTGCGGCTGGCGGGGNTGGC 289

RESULT 2
ID HS1187494 standard; RNA; EST; 226 BP.
AC AA282971;
NI 1925885
DT 05-APR-1997 (Rel. 51, Created)
DT 20-JUN-1997 (Rel. 52, Last updated, Version 3)
DE zt14a08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone 713078 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN [1]
RP 1-226
RA NCI-CGAP;
RT "National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
ET Tumor Gene Index";
EL Unpublished.
CC Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email:
CC Robert.Strausberg@nih.gov This clone is available royalty-free
CC through LML ; contact the IMAGE Consortium (info@image.llnl.gov)
CC for further information. Seq primer: -41m13 fwd. ET from Amersham
CC High quality sequence stop: 220.
FH Key
      Location/Qualifiers
```

```
FT source     1..226
/organism="Homo sapiens"
/notes="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD-),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-TGTTACCAATCGAAGTCGGAGCGGCCCTCATTTTTTTTTTTTTTTT-
3']. Double-stranded cDNA was ligated to Eco RI adaptors (
Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT7T3 vector. Library wen
t
through one round of normalization, and was constructed by
Bento Soares and M. Fatima Bonaldo."
/clone="713078"
/clone_lib="NCI CGAP GCB1"
/tissue_type="germinal center B cell"
/lab_host="DH10B"
FT mRNA       complement(<1..>226)
SQ Sequence 226 BP; 42 A; 56 C; 70 G; 58 T; 0 other;

Query Match   6.9%; Score 22; DB 52; Length 226;
Best Local Similarity 70.6%; Pred. No. 1.48e-04;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 126 tccctggtccatctctgtgagcgctggagtgagctctgtgctgct 176
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 243 TCCTTGGCCCTCTATGGAATGAGGCTCGGGTGGCGGGTGGCTCT 293

RESULT 3
LOCUS         AA282971      226 bp      mRNA      EST      13-JUN-1997
DEFINITION    zt14a08.s1 NCI_CGAP_GCB1 Homo sapiens cDNA clone 713078 3'.
ACCESSION     AA282971
NID           g1925885
KEYWORDS      EST.
SOURCE        human.
ORGANISM      Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae;
Homo.
REFERENCE     1 (bases 1 to 226)
AUTHORS       NCI-CGAP.
TITLE         National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
              Tumor Gene Index
JOURNAL       Unpublished (1997)
COMMENT       Contact: Robert Strausberg, Ph.D.
              Tel: (301) 496-1550
              Email: Robert.Strausberg@nih.gov
              This clone is available royalty-free through LML ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: -41m13 fwd. ET from Amersham
              High quality sequence stop: 220.
              Location/Qualifiers
                source
                  1..226
                    /organism="Homo sapiens"
                    /notes="Vector: p7T3D-Pac (Pharmacia) with a modified
                    polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
                    was prepared from human tonsillar cells enriched for
                    germinal center B cells by flow sorting (CD20+, IgD-),
                    provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
```

I and cloned into the Not I and Eco RI sites of the modified p7r3 vector. Library went through one round of normalization."

```
/clone="898899"  
/clone_lib="NCI_CGAP_Co3"  
/sex="pooled"  
/tissue_type="colon"  
/lab_host="DH10B"  
<1..350  
a 96 c 128 g 55 t
```

Query Match 6.9%; Score 22; DB 27; Length 350;
Best Local Similarity 70.6%; Pred. No. 1.48e-04;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Caps 0;

Db 110 tccctgggtccatctcgtgaggacggctggagtggggcctctgtggtcgt 160
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QV 243 TCCTTGGCCCTCTATGCGAATGAGGCTGCGGTGGCGGGNTGCCTCT 293

RESULT	5	standard; RNA; EST; 350 BP.
ID	HS1271527	
AC	AA480207;	
NI	q2208358	
DT	23-JUN-1997 (Rel. 52, Created)	
DT	23-JUN-1997 (Rel. 52, Last updated, Version 1)	
DE	ne330h02.s1 NCI_CGAP_Co3 Homo sapiens cDNA clone 898899.	
KW	EST.	
OS	Homo sapiens (human)	
OC	Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;	
OC	Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.	

RT "National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
RT Tumor Gene Index";
RL Unpublished.

CC Contact: Robert Straubeberg, Ph.D. Tel: (301) 496-1550 Email:
CC Robert.Straubeberg@nih.gov Tissue Procurement: Elias Campo, M.D.,
CC Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: M.
CC Bento Soares, Ph.D. cDNA Library Arraying: Greg Lennon, Ph.D. DNA
CC Sequencing by: Washington University Genome Sequencing Center Clone
CC distribution: NCI-CCAP clone distribution information can be found
CC through the I.M.A.G.E. Consortium/ILMIL at:
CC www-bis.lnlnl.gov/bbrp/image/image.html Seg primer: -41ml3 fwd. ET
CC from Amersham High quality sequence atop: 345.
FH Key Location/Qualifiers

```

1..350
/organism="Homo sapiens"
/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA
was prepared from 12 pooled bulk tumor samples and primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not I
and cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization."
```

```

FT /clone= 090059
FT /clone_lib="NCI_CGAP_Co3"
FT /sex="pooled"
FT /tissue type="colon"
FT /lab host="DH10B"

```



```
FT mRNA <1..>350
SQ Sequence 350 BP; 71 A; 96 C; 128 G; 55 T; 0 other;

Query Match 6.9%; Score 22; DB 64; Length 350;
Best Local Similarity 70.6%; Pred. No. 1.48e-04;
Matches 36; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

Db 110 tccctgggtccatctctggtgagcgtggtggagtgctctgtgctgct 160
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 243 TCCTTGGCCCTCTATGCAATGAGCGCTGCGGCTGGCGGGTGCTCTCT 293

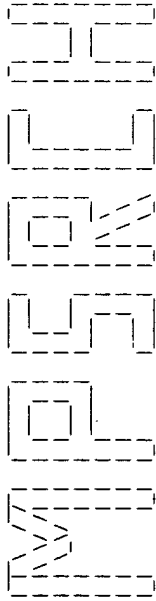
RESULT 6
ID HSA54719 standard; RNA; EST; 458 BP.
AC AM54719;
NI g2177495
DT 07-JUN-1997 (Rel. 52, Created)
DT 07-JUN-1997 (Rel. 52, Last updated, Version 1)
DE zx77b02.s1 Soares ovary tumor NbHOT Homo sapiens cDNA clone 809739
DE 3'.
KW EST.
OS Homo sapiens (human)
OC Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
OC Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
RN [1]
RP 1-458
RA Hillier L., Allen M., Bowles L., Dubuque T., Geisel G., Jost S.,
RA Kucaba T., Lacy M., Le N., Lennon G., Marra M., Martin J.,
RA Moore B., Schellenberg K., Steptoe M., Tan F., Theising B.,
RA White Y., Wylie T., Waterston R., Wilson R.;
RT "WashU-Merck EST Project 1997";
RL Unpublished.
CC Contact: Wilson RK WashU-Merck EST Project Washington University
CC School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis,
CC MO 63108 Tel: 314 286 1800 Fax: 314 286 1810 Email:
CC est@watson.wustl.edu This clone is available royalty-free through
CC LNL; contact the IMAGE Consortium (info@image.llnl.gov) for
CC further information. Seq primer: -41m13 fwd. ET from Amersham High
CC quality sequence stop: 406.
FH Key Location/Qualifiers
FT 1..458
FT /organism="Homo sapiens"
FT /note="Organ: ovary; Vector: pT7T3D (Pharmacia) with a
FT modified polylinker; Site 1: Not 1; Site 2: Eco RI; 1st
FT strand cDNA was primed with a Not I - oligo(dT) primer [5'
FT TGTACCAATCTGAAGTCGAGCGCGCGGCTTTTTTTTTTTT 3'],
FT double-stranded cDNAs size selected, ligated to Eco RI
FT adapters (Pharmacia), digested with Not I and cloned into
FT the Not I and Eco RI sites of a modified pT73 vector
FT (Pharmacia). Library constructed by Bento Soares and
FT M.Fatima Bonaldo."
FT /clone="809739"
FT /clone_lib="Soares ovary tumor NbHOT"
FT /sex="Female"
FT /tissue_type="ovarian tumor"
FT /lab_host="DH10B (ampicillin resistant)"
FT complement(<1...>458)
SQ Sequence 458 BP; 93 A; 126 C; 145 G; 94 T; 0 other;

Query Match 6.9%; Score 22; DB 78; Length 458;
Best Local Similarity 70.5%; Pred. No. 1.48e-04;
Matches 31; Conservative 2; Mismatches 11; Indels 0; Gaps 0;
```

```
Db 165 cgggctgagaccgcctcctgcctgatgcagccaggattgggg 208
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Cp 238 C3GGCTGAGCCGAGCACTGCCTCGGCCGRCGAGCCTTGGGG 195

RESULT 7
LOCUS G25421 294 bp DNA STS 31-MAY-1996
DEFINITION human STS EST405749.
ACCESSION G25421
NID g1347653
KEYWORDS STS sequence; primer; sequence tagged site.
SOURCE human STSs derived from sequences in dbEST and the Unigene
collection.
ORGANISM Homo sapiens
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 294)
AUTHORS Hudson, T.
TITLE Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSs
JOURNAL Unpublished (1995)
COMMENT
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu
Primer A: ACAAAATATTACCAACCCGAAA
Primer B: TTTTATGTACCAATTGCTGAATG
STS size: 150
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul
Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 9.3
Derived from dbEST (genbank accession H81577).
FEATURES
source
Location/Qualifiers
1..294
/organism="Homo sapiens"
/map="48.5 cR from top of Chr14 linkage group"
STS
1..150
primer_bind 1..22
primer_bind complement(127..150)
BASE COUNT 90 a 49 c 48 g 107 t
ORIGIN
```


Cp		234		CTGAGCCAGGACCTGCGCCTCGGRCGRC		206	
RESULT		10					
LOCUS		G29737		419 bp		DNA	
DEFINITION		human STS SHCC-35139.				STS	
ACCESSION		G29737				04-OCT-1996	
NID		g1593288					
KEYWORDS		STS sequence; primer; sequence tagged site.					
SOURCE		human.					
ORGANISM		Homo sapiens					
REFERENCE		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;					
AUTHORS		Vertebrata; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
JOURNAL		Myers, R.M.					
COMMENT		Unpublished (1996)					
		Contact: Richard M. Myers					
		Stanford Human Genome Center (SHCC)					
		Stanford University School of Medicine					
		Department of Genetics, M-344, Stanford, CA 94305, USA					
		Tel: 4157259687					
		Fax: 4157259689					
		Email: myers@shgc.stanford.edu					
		Primer A: ACAGAAATATACACCGGAA					
		Primer B: TTTTGTATGACCAATGCTGAATG					
		STS size: 150					
		PCR Profile:					
		Initial incubation: 94 degrees C for 90 seconds					
		Denaturation: 94 degrees C for 15 seconds					
		Annealing: 62 degrees C for 23 seconds					
		Polymerization: 72 degrees C for 30 seconds					
		PCR Cycles: 30					
		Thermal Cycler: Perkin Elmer 9600					
		Protocol:					
		Template: 25 ng					
		Primer: each 1 uM					
		dNTPs: each 200 uM					
		Taq Polymerase: 0.05 units/ul					
		Total Vol: 10 ul					
		Buffer:					
		MgCl2: 2.5 mM					
		KCl: 50 mM					
		Tris-HCl: 20 mM					
		pH: 8.3					
		Prepared with primer pairs provided by Sandoz, derived from H81577					
		-- Washington University/Merck EST sequence.					
FEATURES		Location/Qualifiers					
source		1..419					
		/organism="Homo sapiens"					
		/map="14"					
STS		1..150					
primer_bind		1..22					
primer_bind		complement(127..150)					
BASE COUNT		121 a 70 c 62 g 164 t 2 others					
ORIGIN							
Query Match		6.6%; Score 21; DB 5; Length 419;					
Best Local Similarity		61.5%; Pred. No. 1.80e-03;					
Matches		24; Conservative 1; Mismatches 4; Indels 0; Gaps 0;					



(TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.
Distribution rights by IntelliGenetics, Inc.

MPsrch pp protein - protein database search, using Smith-Waterman algorithm

```
Run on:      Mon Feb 23 10:59:47 1998;  MasPar time 6.00 Seconds
           252.111 Million cell updates/sec
Tabular output not generated.
```

```
>US-08-836-075A-2
Description: (1-109) from US08836075A.pep
Perfect Score: 781
Sequence: 1 MSTNPKPQRTKNTNRPX.....
```

Sequence: 1 MSTNPKPQRTKRNTPRPX.....GCGWAXWLLSPGSRPNWGP 109

Scoring table: PAM 150
Gap 11

Searched: 111726 seqs, 13889129 residues

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database: a-geneseq30

1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7
8:part8 9:part9 10:part10 11:part11 12:part12 13:part13
14:part14 15:part15 16:part16 17:part17 18:part18
19:part19 20:part20 21:part21 22:part22 23:part23

Statistics: Mean 28.686: Variance 128.255: scale 0.224

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query			DB	ID	Description	Pred. No.
	No.	Score	Match Length				
1	781	100.0	109	19	R96526	Hepatitis C virus typ	8.90e-62
2	777	99.5	117	19	R96539	Hepatitis C virus typ	2.12e-61
3	777	99.5	191	17	R92976	Hepatitis C virus iso	2.12e-61
4	777	99.5	191	17	R92973	Hepatitis C virus iso	2.12e-61
5	777	99.5	502	12	R67591	Hepatitis C virus SR0	2.12e-61
6	777	99.5	502	12	R67589	Hepatitis C virus YS1	2.12e-61
7	777	99.5	3011	12	R67588	Hepatitis C virus gen	2.12e-61
8	775	99.2	191	17	R92977	Hepatitis C virus iso	3.27e-61
9	775	99.2	191	17	R92937	Hepatitis C virus iso	3.27e-61
10	775	99.2	191	17	R92939	Hepatitis C virus iso	3.27e-61
11	775	99.2	191	17	R92941	Hepatitis C virus iso	3.27e-61

ALIGNMENTS

RESULT

ID	R96526	standard:	peptide: 109 AA.
----	--------	-----------	------------------

AC R96526:

DT 07-MAR-1997 (first entry)

DE Hepatitis C virus type 1d isolate BNL1 amino acids 1-103.

KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;

KW PCR; primer; probe; antibody; infection.

OS Hepatitis C virus.

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT Misc difference 20

FT /label= His, Gln

FT Misc difference 21

FT /label= Gly

FT /note= "amino acid in this position is designated X in

FT the specification, but codon usage shows that the

FT only possible amino

FT Misc_difference 22

FT /label= Val

FT /note="amino acid in this position is designated X in

FT the specification, but codon usage shows that

FT only possible amino

FT Misc_difference 23

FT /label= any amino a

FT Misc_difference 24

FT /label= any amino a

FT Misc_difference 44

Feb 23 10:57

/home/fuller/feb1098/US-08-836-075A-2.rag

3

```

FT /label= any amino acid
FT Misc difference 69
FT /label= Arg
FT /note= "amino acid in this position is designated X in
FT the specification, but codon usage shows that the
FT only possible amino acid at this pos. is Arg"
FT Misc difference 71
FT /label= Ala, Pro
FT Misc difference 95
FT /label= Gly
FT /note= "amino acid in this position is designated X in
FT the specification, but codon usage shows that the
FT only possible amino acid at this pos. is Gly"
PN W09613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI; 96-251460/25.
DR N-PSDB; T27937.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 25; Fig 3; 150pp; English.
CC The sequences R96526-R96578 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/EL, NS4 or NS5B regions of the genome.
CC This sequence represents amino acids 1-103 from the HCV type 1d isolate
CC BNL7.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/EL and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to generate the peptides R96424-R96524. The
CC sequences can also be used to generate the peptides R96424-R96524. The
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 109 AA;

Query Match 100.0%; Score 781; DB 19; Length 109;
Best Local Similarity 100.0%; Pred. No. 8.90e-62;
Matches 109; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 mstnpkqrktrntnrpxxxxpgggqivggvylprgrgxvratrktsersqprg 60
    |||||||
Qy 1 MSTNPKQRKTRNTNRPRXXXXPGGGQIVGGVYLPRGRGXVTRATRKTSERSQPRG 60

Db 61 rrgpikaxregswagqgypwplynecgcwaxwllsprgtrpnwqp 109
    |||||||
Qy 61 RRQIPKAXRKGESWAGQGYPPWPLYNCGCGWAXWLLSPRGSRPNWGP 109

RESULT 2
ID R96539 standard; peptide; 117 AA.
AC R96539;
DT 02-OCT-1996 (first entry)
DE Hepatitis C virus type 4k isolated BNL7 amino acids 1-117.
KW Hepatitis C virus; subtype; polymerase chain reaction; amplification;

```

Feb 23 10:57

/home/fuller/feb1098/US-08-836-075A-2.rag

4

```

KW PCR; primer; probe; antibody; infection.
OS Hepatitis C virus.
FH Key Location/Qualifiers
FT /note= "any amino acid"
PN W09613590-A2.
PD 09-MAY-1996.
PF 23-OCT-1995; E04155.
PR 21-OCT-1994; EP-870166.
PR 28-JUN-1995; EP-870076.
PA (INNO-) INNOGENETICS NV.
PI Maertens G, Stuyver L;
DR WPI; 96-251460/25.
DR N-PSDB; T27950.
PT Hepatitis C virus poly:nucleic acid unique to unidentified sub:type
PT - used to develop probes and primers for new sub:types and vaccines
PT to prevent and treat infection
PS Claim 25; Fig 3; 150pp; English.
CC The sequences R96526-R96578 represent novel sequences isolated from
CC hepatitis C virus subtypes different from subtypes 1a-c, 2a-d, 3a-f,
CC 4a-j, 5a and 6a. They esp. from the novel subtypes 1d-f, 2e-i, 2k, 2l,
CC 3g, 4k-m, 7a-c or types 9, 10 or 11. The sequences corresp. to the 5'
CC untranslated region (UR), the Core/EL, NS4 or NS5B regions of the genome.
CC This sequence represents amino acids 1-117 from the HCV type 4k isolate
CC BNL7.
CC The new HCV types were isolated from patients with chronic HCV from the
CC Benelux countries, France, Cameroon and Vietnam, because of their
CC aberrant reactivities. The RNA was extracted, cDNA synthesised and PCR
CC amplified, cloned and genotyped. The 5'UR, Core/EL and NS5B regions were
CC sequenced either directly or partially and used to classify the new
CC viruses into (sub)types based on comparison with known sequences.
CC The sequences were used to generate the peptides R96424-R96524. The
CC sequences can also be used to generate the peptides R96424-R96524. The
CC sequences can also be used to generate the peptides R96424-R96524. The
CC detection of HCV in a sample. The polypeptides can be used to detect
CC anti-HCV antibodies, for HCV typing or to prevent HCV infections.
SQ Sequence 117 AA;

Query Match 99.5%; Score 777; DB 19; Length 117;
Best Local Similarity 91.7%; Pred. No. 2.12e-61;
Matches 100; Conservative 1; Mismatches 8; Indels 0; Gaps 0;

Db 1 mstnpkqrktrntnrpxxxxpgggqivggvylprgrgxvratrktsersqprg 60
    |||||||
Qy 1 MSTNPKQRKTRNTNRPRXXXXPGGGQIVGGVYLPRGRGXVTRATRKTSERSQPRG 60

Db 61 rrgpikaxregswagqgypwplynecgcwaxwllsprgtrpnwqp 109
    |||||||
Qy 61 RRQIPKAXRKGESWAGQGYPPWPLYNCGCGWAXWLLSPRGSRPNWGP 109

RESULT 3
ID R92976 standard; Protein; 191 AA.
AC R92976;
DT 02-OCT-1996 (first entry)
DE Hepatitis C virus isolate 26 core protein.
KW HCV; EL; envelope 1; core protein; HCV genotyping; antibody; vaccine;
KW hepatitis.
OS Hepatitis C virus.
PN W09605315-A2.
PD 22-FEB-1996.
PF 15-AUG-1995; U10398.
PR 15-AUG-1994; US-290665.
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (USSH ) US SEC DEPT HEALTH.

```


Query Match 99.2%; Score 775; DB 17; Length 191;
 Best Local Similarity 89.9%; Pred. No. 3.27e-61;
 Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnkpqrktkntnrpqdvkfpqggqivgvyllprgrglvratrktsersqrg 60
 Qy 1 MSTNPKPQRKTNTNRPPXXXXPFGGQIVGGVYLLPRGRGXGVRATRKTSERSQRG 60

Db 61 rrcpikarrpegrtwagpywpvlyngcgwagwllsprgsrpswgp 109
 Qy 61 RRQIPKAXRXEGRSWAQPGYPWPVLYNGCGWAXWLLSPRGRPNWGP 109

RESULT 13

ID R44010 standard; protein; 191 AA.
 AC R44010;
 DT 12-MAY-1994 (first entry)
 DE Hepatitis C virus core protein.
 KW HCV; non-A, non-B hepatitis virus; NANBHV; immunodominant region;
 KW capsid protein; vaccine; antigen.
 OS Hepatitis C virus.
 FH Key Location/Qualifiers
 FT Peptide 1..45
 FT /label= S42G
 FT /note= "peptide 1"
 FT Peptide 38..81
 FT /label= p42Y
 FT /note= "peptide 2"
 FT Peptide 77..115
 FT /label= R40R
 FT /note= "peptide 3"
 FT Region 1..45
 FT /label= immunodominant region
 FT /note= "novel antigenic peptides contain sequences from this region"
 PN EP-569309-A.
 PD 10-NOV-1993.
 PF 06-MAY-1993; 420183.
 PR 06-MAY-1992; FR-005763.
 PA (DALB/) DALBON P.
 PA (INMR) BIO MERIEUX.
 PI Dalbon P, Jolivet M;
 DR WPI; 93-353383/45.
 PT New antigenic polypeptide(s) from hepatitis C virus - derived from N-terminal region of core protein, and related antibodies, useful in diagnosis, vaccination and treatment
 PS Claim 1: Fig 1; 24pp; French.
 CC Of peptides 1-3 (see Features Table), peptide 1 was found to be from the immunodominant region of HCV core protein. Further analysis of the N-terminal 45 amino acids of the HCV core protein showed that some epitopes are present in the first 21 amino acids, some in 22-45 and some at the junction of these two subregions. Preferred antigenic CC peptides are derived from S42G; they are useful for detecting HCV CC antibodies and in vaccines against HCV.
 SQ Sequence 191 AA;

Query Match 99.2%; Score 775; DB 8; Length 191;
 Best Local Similarity 89.9%; Pred. No. 3.27e-61;
 Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnkpqrktkntnrpqdvkfpqggqivgvyllprgrglvratrktsersqrg 60
 Qy 1 MSTNPKPQRKTNTNRPPXXXXPFGGQIVGGVYLLPRGRGXGVRATRKTSERSQRG 60

Db 61 rrcpikarrpegrtwagpywpvlyngcgwagwllsprgsrpswgp 109
 Qy 61 RRQIPKAXRXEGRSWAQPGYPWPVLYNGCGWAXWLLSPRGRPNWGP 109

RESULT 14

ID R92938 standard; Protein; 191 AA.
 AC R92938;
 DT 02-OCT-1996 (first entry)
 DE Hepatitis C virus isolate S14 core protein.
 KW HCV; E1; envelope 1; core protein; HCV genotyping; antibody; vaccine;
 KW hepatitis.
 OS Hepatitis C virus.
 PN W09605315-A2.
 PD 22-FEB-1996.
 PF 15-AUG-1995; J10398.
 PR 15-AUG-1994; US-290665.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA (USSH) US SEC DEPT HEALTH.
 PI Bukh J, Miller RH, Purcell RH;
 DR WPI; 96-139709/14.
 DR N-PSDB; T16612.
 PT DNA and amino acid sequence of HCV envelope 1 and core proteins - used to determine HCV genotype and as vaccines against HCV infection
 PS Claim 4; Page 182; 340pp; English.
 CC R92936-R92987 are HCV core proteins derived from 52 different HCV isolates. Isolated cDNA sequences are used for the prodn. of primers useful for detecting the presence of HCV in a sample, the primers CC are also useful for HCV genotyping. Proteins encoded by the cDNAs CC can be used in vaccines for immunising against HCV infection. The CC proteins may also be used to detect antibodies against HCV in serum, CC saliva, lymphocytes or other mononuclear cells. The antibodies may CC be used in the prevention of HCV infection.
 SQ Sequence 191 AA;

Query Match 99.2%; Score 775; DB 17; Length 191;
 Best Local Similarity 89.9%; Pred. No. 3.27e-61;
 Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnkpqrktkntnrpqdvkfpqggqivgvyllprgrglvratrktsersqrg 60
 Qy 1 MSTNPKPQRKTNTNRPPXXXXPFGGQIVGGVYLLPRGRGXGVRATRKTSERSQRG 60

Db 61 rrcpikarrpegrtwagpywpvlyngcgwagwllsprgsrpswgp 109
 Qy 61 RRQIPKAXRXEGRSWAQPGYPWPVLYNGCGWAXWLLSPRGRPNWGP 109

RESULT 15

ID R79222 standard; Protein; 967 AA.
 AC R79222;
 DT 08-DEC-1995 (first entry)
 DE pHCV141-encoded sequence.
 KW pHCV141; amyloid precursor protein; APP; hepatitis C virus; HCV; E1;
 KW E2; fusion protein; HEK-293; vaccine; vector; prc/CNV;
 KW protein secretion; glycosylation.
 OS Synthetic.
 PN W09520664-A.
 PD 03-AUG-1995.
 PF 27-JAN-1995; U01087.
 PR 28-JAN-1994; US-188281.
 PA (ABBO) ABBOTT LAB.
 PI Desai SM, Devare SG, Watanabe S, Yamaguchi J;

DR WPI; 95-275449/36.
PT New mammalian expression systems for HCV proteins - express fusion
PT proteins comprising amyloid precursor protein and HCV E1 and/or E2
PT protein
PS Disclosure; Page 59-62; 89pp; English.
CC RNA from serum or plasma of a chimpanzee infected with HCV was
CC converted to cDNA and PCR amplified using primers based on HCV
CC sequences. 7 Adjacent HCV DNA fragments were generated which
CC together encoded the HCV sequence given in R79232. Fragments
CC from 2 clones, pHCV141 and pHCV150 (see R79223), were combined to
CC generate pHCV176 (see R79221). This was used to construct
CC APP-HCV-E1+E2 fusion proteins that were glycosylated and secreted
CC from HEK-293 transfectants.
SQ Sequence 967 AA;

Query Match 99.2%; Score 775; DB 13; Length 967;
Best Local Similarity 89.9%; Pred. No. 3.27e-61;
Matches 98; Conservative 2; Mismatches 9; Indels 0; Gaps 0;

Db 1 mstnkpqrktkntnrpqdkvfgggqivggvllprgrlgvratrkttsersqrg 60
|||||
Qy 1 MSTNKPQRKTNTNRPPXXXXXPGGGQIVGGVYLLPRGRCYGVATRTKTSERSQPRG 60
|||||
Db 61 rrqpiqkrrpgrtwagqypwplynecqwgagllsprqsrpewqp 109
|||||
Qy 61 RRQPIPKAXRXEGRSWAQPGFPHLYGEGCGXAXWLLSPRGSRPNWGP 109
|||||

Search completed: Mon Feb 23 11:00:48 1998
Job time : 61 secs.